SUMMARY FUNCTIONS FOR DATA IN THE PRESENCE OF COMPETING RISKS

by

Yongyun Zhao

B.S., Qingdao University of Science and Technology, 1992

M.S., Qingdao University of Science and Technology, 1997

Submitted to the Graduate Faculty of

the Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2008

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This dissertation was presented

by

Yongyun Zhao

It was defended on

July 31, 2008

and approved by

Dissertation Advisor:
(Joyce) Chung-Chou H. Chang, PhD,
Associate Professor, Department of Medicine and Biostatistics,
School of Medicine, University of Pittsburgh

Co-Advisor:
Lisa A. Weissfeld, PhD,
Professor, Department of Biostatistics,
Graduate School of Public Health, University of Pittsburgh

Committee Member:
Jong-Hyeon Jeong, PhD,
Associate Professor, Department of Biostatistics,
Graduate School of Public Health, University of Pittsburgh

Committee Member:
Howard E. Rockette, PhD,
Professor, Department of Biostatistics,
Graduate School of Public Health, University of Pittsburgh

Committee Member:
Mark S. Roberts, MD, MPP,
Professor, Department of Medicine,
School of Medicine, University of Pittsburgh

SUMMARY FUNCTIONS FOR DATA IN THE PRESENCE OF COMPETING RISKS

Yongyun Zhao, PhD

University of Pittsburgh, 2008

Although cumulative incidence function (CIF) estimates are commonly used to describe the failure probabilities when competing risks are present, the CIF has limitations in some scenarios. The objective of our research was to propose new summary functions or modify CIF to overcome the limitations.

In observational studies or nonrandomized trials, CIF estimates can be biased if the distribution of a confounding variable differs among treatment groups. To reduce the bias, we developed an adjusted CIF (ACIF) estimator that is based on the use of inverse probability weighting. We derived the estimation and inference procedures, and then used simulation studies to evaluate the performance. To illustrate the application of ACIF, we used the example of liver transplant candidates with various types of end-stage liver disease.

We also developed a series of adjusted survival functions to estimate the "net survival probability for a specific outcome (the main event), based on the degree of correlation between this event and the competing events. First, for cases in which there is a perfect negative correlation, we constructed an adjusted survival function that uses the Kaplan-Meier estimator and the inverse probability of censoring weight (IPCW) for patients who experience competing events. Second, for cases in which there are imperfect negative correlations, we constructed an adjusted survival function that uses the combination of the IPCW and the adjusted number at risk under the constraints of the lower and upper bounds. Third, for cases in which there are positive correlations, we constructed an adjusted survival function that uses the adjusted number of main events under the constraints of the lower and upper

bounds. To recover the contribution that the competing events make to the net survival probability, we incorporated auxiliary variables into the adjusted number at risk or the adjusted number of main events. We derive the estimation and inference procedures. We demonstrate the use of adjusted survival functions in data derived from patients who had emphysema, a severe form of chronic obstructive pulmonary disease.

The public significance of this work is to provide new approaches to summarize the survival data with competing risks.

Keywords: adjusted cumulative incidence function, competing risks, cumulative incidence function, inverse probability weighting, inverse probability of censoring.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

I would like to express my sincere gratitude to my advisor, Dr. Chung-Chou H. Chang, for her support and guidance throughout my research. Her perpetual energy and enthusiasm in research had motivated and inspired me. I owe her a lot, from the general statistical methodologies to the detailed writing and presentation skills. Her generous and tireless mentorship made all the difference in my academic career.

I am grateful to my co-advisor, Dr. Lisa A. Weissfeld, for her heartfelt encouragement and constant mentoring for five years. To Dr. Howard E. Rockette, for involving me in his NIH training grant, my experience as a trainee prepared me well for becoming a biostatistician and researcher.

I owe a special thanks to Dr. Mark S. Roberts and NETT study group for providing me the datasets. Also I want to thank Dr. Jong-Hyeon Jeong for taking time to serve in my dissertation committee.

Being part of the department of Biostatistics was a very special and invaluable journey for me. Thanks to all faculty members and staffs for creating such a wonderful environment for learning and working.

Finally, I would like to thank my parents, my husband Jianjun, my son Kevin and my daughter Jessica for their deep love and strong support.

1.0 INTRODUCTION

A competing risk is defined as an event whose occurrence either precludes or affects the occurrence of the other event. In the usual competing risk setting, the endpoint consists of several distinct events of interest and the eventual failure being attributed to one event exclusively of the others. Under this framework, an individual is exposed to several risks simultaneously but can only fail from one of these risks. There is an alternative way to define competing risks, in which an individual may fail from several causes. The current study only focuses on the usual competing risk setting.

When competing risks are present, a summary function is either used to describe the time to event for each cause, without focusing on any specific cause or used to describe the time to event for a specific cause of interest. It is inappropriate to use the complement of the Kaplan-Meier (KM) estimator, 1−KM, to summarize the data because this method censors the competing events after they happen and therefore overestimates the failure probability of the event of interest. Currently, the cumulative incidence function (CIF) is the most widely used summary function to address these two types of questions. However, the CIF also suffers from some limitations. When survival data has an unevenly distributed confounder among groups, the estimated CIF could be biased and misleading. Moreover, the CIF addresses the marginal failure probability of interest and therefore the resulting function only presents partial information if the failure probabilities of the competing events do not present simultaneously. To address these two issues, two new functions are proposed to summarize survival data with competing risks.

1. In the first part of the study, a new method is developed to estimate the failure probability of the event of interest when competing risks are present and when the effects of

confounding variables cannot be ignored. Specifically, instead of using stratification, the proposed estimator is based on the use of inverse probability weighting. After adjusting the unbalanced confounder, the selection bias is removed and an unbiased estimator of the CIF is obtained.

2. In the second part of the study, a series of adjusted survival functions is proposed to estimate the "net" survival probability for a specific outcome (the main event), based on the degree of correlation between this event and the competing events. The new estimators are constructed based on Peterson's bound and the KM estimate. Specifically, we treat the risks with negative correlation and the risks with positive correlation differently. For the negative correlation, by using the adjusted number at risk which incorporates the information recovered from the auxiliary variables for the second event and the inverse censoring weight, a generalized adjusted survival function is constructed based on the net survival bound. For the positive correlation, we introduce the net survival bound first, then a generalized adjusted survival function is constructed by using the adjusted number at death, which also incorporates the recovered information of the second event from auxiliary variables.

Two motivational data sets are described in Chapter 2. Previous research on the summary functions for data with competing risks are presented in Chapter 3. For both of the proposed summary functions, the derivation of the properties, the inference procedure, and the simulation plan for assessing the performance of the proposed functions are described in detail in Chapters 4 and 5. The summary of the current work and the plan for future work are discussed in Chapter 6.

2.0 MOTIVATION DATASETS

2.1 THE LIVER TRANSPLANT DATA FOR THE ADJUSTED CUMULATIVE INCIDENCE FUNCTION

In the United States, there are more patients on the waiting list for a liver transplant than the number of organs available. In order to allocate organs in an efficient and appropriate manner, patients are ordered in the waiting list according to the likelihood that a patient will die while awaiting transplant. For all patients on the waiting list, some of them may die before transplant (pre-transplant death), some of them may receive a liver transplant, some of them may be removed from the list due to other causes (like improvement of disease and drop out), while the rest of them are alive at the study cutoff time. Therefore, if we are interested in estimating the probability of pre-transplant death, receiving a transplant and be removed from the waiting list due to other reasons are competing risks.

Currently, the scores from the model for end-stage liver disease (the MELD scores) are used to rank the order and the priority of a patient for receiving a transplant. The MELD score is an index of the severity of the liver disease. The higher the score is, the more likely the patient will die before receiving a transplant. The MELD score was derived from a Cox proportional hazards regression model. In this model, time to transplant and time to removal from the list due to other reasons were both treated as censoring.

There are several different types of liver disease. Severity of illness varies with the etiology of the liver disease. Therefore, when we compare the probability of pre-transplant death among different types of liver disease, we have to take into account the different severity of illness among these groups. In other words, severity of illness measured by the MELD

score is an unbalanced confounder. The naive CIF estimator cannot be applied without modification because of this selection bias.

2.2 THE EMPHYSEMA DATA FOR THE ADJUSTED SURVIVAL FUNCTION

The National Emphysema Treatment Trials (NETT) is the first multi-center clinical trial designed to determine the role, safety, and effectiveness of bilateral lung volume reduction surgery (LVRS) in the treatment of emphysema. Patients who met the inclusion criteria were randomized into the group of using medical treatment only (N=610) or the group of undergoing surgery in addition to medical treatment (N=608). The trial finished to enroll patients in July 2002 and followed them up until December 2002. The study found that on average patients with severe emphysema who underwent LVRS with medical treatment were more likely to function better, however, the 5-year survival rates between the two groups were comparable. Although there is no significant survival benefit for patients with LVRS who under medical treatment, researchers would like to know whether their health-related quality of life (HRQL) improved significantly compared to those who had medical therapy only. The HRQL was measured over time using the St. George Respiratory Questionnaire. If a patient had dropped 8 points of the HRQL measure from the baseline, this patient's HRQL considered to have improved significantly. The main study question is to estimate the probability of HRQL improvement over the 5-year post-randomization follow up period and to compare the probabilities between the patients assigned to these two treatment groups. In this study, the main outcome of interest is time to HRQL improvement and death before reaching the HRQL improvement is a competing risk. Because those who died had worse HRQL compared to those who lived, death before reaching the HRQL is subject to dependent censoring. A single summary probability function of HRQL improvement is of main interest and this estimated function needs to take the dependent censoring due to the competing risk into account.

In the liver transplant dataset, our main event of interest is death and the patients' transplant can be treated as a competing risk. However, once the patient received the transplant, the risk of death changes. So in this case, death after transplant is not unobservable anymore and the better way to analyze the data is to treat transplant as a time-dependent variable. We can then test the effect of transplant on the overall survival using the Cox proportional hazards model.

3.0 LITERATURE REVIEW

3.1 BASIC CONCEPTS OF COMPETING RISKS

A competing risk is defined as an event whose occurrence either precludes or affects the occurrence of the other event. In the usual competing risk situation, the endpoints consist of several distinct events of interest and the eventual failure is attributed to one event exclusive of the others. Here are two main examples for competing risks:

1. When relapse of a specific disease (such as leukemia, cancer) is the primary event of interest, deaths from other causes (like aging) without relapse of the disease are competing risks.
2. Another classical example is cause-specific mortality, such as death from heart disease, death from cancer, death from other causes, etc.

Under the usual competing risk framework, a subject can only fail from one of the many distinct causes. There is an alternative way to define the competing risk situation when a subject may fail from several causes. In this paper, our study only focuses on the usual competing risk setting. In the competing risk setting, a subject is exposed to several risks simultaneously, and he can either fail from one of the causes or be censored from the study. So if we have one primary event of interest, we call it the "interest event", the other risks are then the "competing risks".

When competing risks are present, the time-to-event data are complicated because several distinct risks are acting on a subject at the same time, and the associations among these risks are not identifiable. Of course, investigators want to answer different questions related to various research areas. In general, there are two types of situations. One situation is to

describe the time-to-event change over time for each cause, without focusing on any specific cause. The other situation is to concentrate on one interest cause, and try to answer questions around this "interest cause".

3.2 ASSUMPTION OF INDEPENDENT CENSORING

To understand the censoring mechanism is important in analyzing survival data. There are many different types of censoring and the most common one is the right censoring, which refers to censoring that always occurs on the right side of the time continuum. For individuals with right censoring, we only know that they are still alive before a certain time, but no information is available about what will happen after that time. In survival data with competing risks, "right censoring" may alter the hazards of the interest events. Specifically, if the patients drop out the study in a non-random pattern due to severity of illness, then the patients remaining in the risk set may be quite different from the original study population. In this case, we have "dependent censoring" and the censoring mechanism does affect the analysis results. However, standard methods assume "independent censoring" meaning that

$$\lim_{\Delta \to 0} \frac{P(T_i \leq t + \Delta, R = r | T_i \geq t)}{\Delta} = \lim_{\Delta \to 0} \frac{P(T_i \leq t + \Delta, R = r | T_i \geq t, C_i \geq t)}{\Delta},$$

where $R = r$ denotes the failure from cause r. Under this model: all cause specific hazards are independent from whether a subject is censored or not. We need independent censoring to obtain unbiased and valid estimates for censored survival data.

3.3 CURRENT SUMMARY FUNCTIONS

In general, there are three ways to summarize these data:

1. Analyze the interest event only by treating the events from competing risks as censored (independent censoring).
2. Analyze one joint endpoint combined by the interest event and competing events.

3. Analyze the competing risks using the cumulative incidence function.

The first approach is not appropriate because competing risks are dependent with the interest event, which is known as "dependent censoring". The second approach is correct, but it is not sufficient to address the important aspects of research questions. The third approach is preferable, but it has some limitations. First, the cumulative incidence function of the interest event and competing risks should be considered simultaneously in order to address the failure probability of interest event correctly. Second, the cumulative incidence function can be biased and misleading if we have an unevenly distributed confounder involved among groups.

3.3.1 Cause-Specific Hazard

For competing risks problems, we have two different tools to describe the data: the failure probability and the hazard rate. The hazard function is of the form:

$$\lambda^r(t) = \lim_{\Delta \to 0} \frac{P(t \leq T < t + \Delta, R = r | T \geq t)}{\Delta}$$

For estimation purposes, we can also define it as:

$$\lambda^r(t) = P(T = t, R = r | T \geq t).$$

This is the cause specific hazard of cause r at time t. We can estimate this quantity using the number of events of cause r divided by the number at risk. The semi-parametric regression models for competing risks focus on modeling the "cause specific hazard", and a lot of work has been done in this area.

There are two main disadvantages of the cause specific hazard. First, the hazard is not intuitively interpretable for the physicians. For competing risks problems, we are usually interested in summarizing the likelihood of the occurrence of a particular competing risk instead of the hazard rate. The hazard functions describe the change rate of the failure probability, they do not quantify the ultimate benefit of the treatment to the patient. Second, the hazard functions, like the density functions, are difficult to estimate accurately (Pepe and Mori 1993). And in many cases, the effect of covariates on the failure probability is

quite different from the effect on the cause specific hazard. Therefore, the estimation of the failure probabilities becomes our primary goal in presence of the competing risks in most cases.

3.3.2 Cause-Specific Failure Probability

Klein and Moeschberger (2003) summarized failure probability into three types of probabilities: crude probability, net probability and partial probability.

Crude probability, also called the cause-specific subdistribution function and cumulative incidence function (CIF), is defined as

$$I^r(t) = P[T \leq t, R = r] = \int_0^t \lambda^r(t) exp\{-\Lambda_T(u)\} du,$$

where $\Lambda_T(t) = \sum_{r=1}^{M} \int_0^t \lambda^r(u) du$. The cumulative incidence function is the probability of death from a particular cause in the real world where all other risks are acting on the individual. This is the most acceptable non-parametric estimator in the presence of competing risks. $I^r(t)$ is a function of hazards of all failure types and is estimable directly without making any assumption of the joint distribution of the failure times. Also, $I^r(t)$ is a monotone increasing function with $I^r(\infty) < 1$, so that it is not a true distribution function, and referred to as a "subdistribution function". Methods of estimating the probability of failure are discussed in many papers (Dinse and Larson, 1986; Gaynor et al., 1993; Pepe and Mori, 1993; Klein and Moeschberger, 2003) and the inferences of CIF have also been worked out in this setting (Gray, 1988; Pepe, 1991). Satagopan et al. (2004) gave five steps to calculate the CIF estimator, they also illustrated nonparametric estimation of CIFs using two published datasets and compared them with the Kaplan-Meier approach.

CIFs have several advantages. First, they are the unbiased estimators for the probability of event of interest cause. Second, CIFs are calculated directly from the observed data without making any assumption about the potential failure times, In other words, CIFs make no assumption about the association of the interest risk and the competing risks. Third, CIF estimates are non-parametric estimates, and easy to calculate. However, CIFs still have some limitations. First, the competing causes of failure are not jointly treated,

that is, a separate model is fitted for each failure type, treating other failures as censors. This therefore does not allow for a direct comparison of parameter estimates corresponding to the various failure types. Second, CIFs are only valid for reference under the same kind of population with the same setting of competing risks. If the competing risks change with the different situation, the CIFs are not comparable. Third, the CIF is not a true marginal distribution, it is only a cause specific failure probability when several causes are acting together.

The net probability is interpreted as "the death as a result of the cause of interest if the other competing risks could be removed". The net probability, $S_i(t)$, can be derived as a marginal distribution from the joint distribution while taking $t_j = 0$ for all $j \neq i$. In the case when only one independent competing risk exists, if we treat the competing risk as random censoring, the net probability can also expressed as a function of the crude probability. Furthermore, Peterson (1976), Klein (1988), Zheng and Klein (1995) showed that the net probability is bounded by (1 - the crude probability) and discussed the possible tightness of bounds under some dependence structures of the joint distribution. The partial crude probability is the probability of death in the hypothetical world where some competing risks could be removed. This estimator is similar to the net probability and is barely discussed in papers.

It is well known that the Kaplan-Meier method is inappropriate for estimating the failure probability in presence of competing risks (Pepe and Mori, 1993; Klein and Gooley, 1999). The Kaplan-Meier estimator (Kaplan and Meier 1958), is called the Product-Limit estimator. This method is widely applied for time-event data with only one endpoint. The censoring mechanism for the event is restricted to be independent, which means the potential censoring time is unrelated to the survival outcome. The cause specific failure probability can be estimated by $1-\text{KM}$, where KM is the Kaplan-Meier estimator for the survival function of the cause of interest. It is well known that this is actually a "pure" probability based on a very strong assumption. When the patient fails at time T, we treat him as being censored at time T. In reality, this assumption seems too optimistic and untestable.

Gooley et al (1999) pointed out that the censored observations who failed from other causes led to the biasness of $1-\text{KM}$ and the non-interpretability of the estimator. In practice,

patients failing from competing risks are treated as censored at the time of failure. In the competing risks setting, the number of people at risk is reduced by the failure from competing risks. So the failure probability of interest event should depend on both the hazard rates of the event of interest and the hazard rates of competing risks. While 1−KM is only the hazard function of interest event, it overestimates the failure probability. If the competing risks do not exist, then 1−KM is equal to CIF.

Tai B.C et al (2001) extended the Kaplan-Meier method for the competing risk setting, where a patient may experience more than one event during the study. They introduced two ways to handle the data, "ignore" and "censor" methods. For considering failure probability of type l, in the KM ignore method, all events of type l are included while ignoring all other non-type l events; in the KM censor method, only the first event of type l is included, and all non-type l events are censored. We can see that these estimators do not fit in the usual competing risk frameworks and can not compare with the CIFs.

Pepe and Mori (1993) developed conditional probability estimators, which can be applied for summarizing failure time data. Let R denote the event of interest, and $\overline{DR}$ denote the failure from the remaining events. An estimator of the conditional probability has the form:

$$C\hat{P}_R(t) = \frac{\hat{pr}(\text{had interest event by t})}{1 - \hat{pr}(\text{did not have non-interest events by t})} = \frac{\hat{P}_R(t)}{1 - \hat{P}_{\overline{DR}}(t)},$$

Where,

$$\hat{P}_R(t) = \int_0^t \hat{S}^-(u)\frac{dN_R(u)}{Y(u)}, \qquad \hat{P}_{\overline{DR}}(t) = \int_0^t \hat{S}^-(u)\frac{dN_{\overline{DR}}(u)}{Y(u)}.$$

The authors also discussed the large sample properties of this estimator and presented two-sample test statistics. Actually this new estimator is a function of the CIFs. It is a monotone increasing function and can be interpreted as the proportion of patients who had the event of interest among those who survive from the remaining.

When analyzing survival data with competing risks, we need to notice that: given we observe (T, R), we can never distinguish dependent competing risks from a pair of independent competing risks. This is referred to as the "identifiability dilemma" by Klein and Moeschberger (2003). In other words, we are unable to analyze associations or relationships among different risks. Crowder (1994) reviewed problems of "identifiability" in competing

risks. He also discussed the traditional way of modeling competing risks via latent failure times.

3.4 INFERENCE TEST FOR COMPETING RISKS

The log rank test is a test of equality of the cause specific hazard ($\lambda_r(t)$) since the cause specific survival is a simple function of this hazard in the absence of competing risks. When there are competing risks involved, the CIF or other marginal functions are not a simple function of the cause specific hazard any more. So the log rank test is invalid for comparing two CIFs.

Consider the following hypotheses for testing the equality of two CIFs:

$$H_0\text{: } I_1^r(t) = I_2^r(t) \text{ for all } t \leq \tau \text{ and}$$
$$H_a\text{: } I_1^r(t) \neq I_2^r(t) \text{ for some } t \leq \tau,$$

where τ is the longest time at which both groups have at least one participant at risk. Gray (1988) proposed a class of generalized linear rank statistics for testing the equality of CIFs. For comparing the CIFs of two groups, the test was based on a score of the form

$$\int_0^\tau W(t)[\{1-\hat{I}_1^r(t-)\}^{-1}d\hat{I}_1^r(t) - \{1-\hat{I}_2^r(t-)\}^{-1}d\hat{I}_2^r(t)],$$

where W(t) is a suitably chosen weight function. Basically, the above test compares the weighted averages of the "sub-distribution hazards", $i_k^r/(1-I_k^r)$.

Pepe (1991) proposed a different class of test statistics, not based on ranks, for comparing rather general functions, such as marginal functions (CIFs) and conditional probability functions. This statistic is the cumulative weighted difference and has the form:

$$WP = (\frac{n^1 n^2}{n^1+n^2})^{1/2}\int_0^\tau \hat{W}(t)\{\hat{P}_1^r(t) - \hat{P}_2^r(t)\}dt.$$

She also showed that, under the null hypothesis of equality, these test statistics are asymptotically normal with mean 0, and derived consistent variance estimators. In the Pepe and

Mori (1993) paper, they chose the weight function

$$\hat{W}(t) = \hat{C}_1^-(t)\hat{C}_2^-(t)/\{\frac{n^1}{n}\hat{C}_1^-(t) + \frac{n^2}{n}\hat{C}_2^-(t)\},$$

where $1 - \hat{C}_1^-(.)$ is the left continuous Kaplan-Meier estimator of the censoring distribution function in the 1st group. This function down-weighs the later time points.

Lin (1997) proposed a resampling technique for developing the approximate distribution of estimator of CIF. Based on the approximate distribution, the author built the confidence bands for CIF and constructed Kolmogorov-Smirnov type tests for comparing two CIF curves,

$$D(t) = W(t)[\{\hat{I}_1^r(t) - \hat{I}_2^r(t)\} - \{I_1^r(t) - I_2^r(t)\}],$$

where $W(t)$ is a weight function. Assuming that the observations from the two groups are independent, use the resampling technique to generate the approximate distributions of $\{\hat{I}_1^r(t) - I_1^r(t)\}$ and $\{\hat{I}_2^r(t) - I_2^r(t)\}$ and then evaluate the distribution of D(.). Since $\{\hat{I}_1^r(t) - \hat{I}_2^r(t)\}$converges to $\{I_1^r(t) - I_2^r(t)\}$, it will be non-zero for some t if H_0 does not hold. Hence the use of the Kolmogorov-Smirnov type statistic

$$Q = sup_t W(t)|\hat{I}_1^r(t) - \hat{I}_2^r(t)|,$$

will yield an omnibus test, consistent against any alternatives under which $I_1^r(t) \neq I_2^r(t)$ for some t within the range of the data.

3.5 SURVIVAL DATA ANALYSIS WITH CONFOUNDERS

A confounder refers to a third variable that can indirectly distort the statistical relationship between two variables under manipulation or observation. When confounding variables are present, the estimation of treatment effect can be biased due to the unbalanced distribution of confounders among groups. For example, in an observational study, we may have more sicker patients in the treatment group when compared to the control group. When we analyze the treatment effect based on the original observations, we may conclude that the treatment is not effective. While it may be due to the selection bias of the patients. Therefore,

adjusting for confounders is essential for these studies, such as in observational studies and non-randomized clinical trials.

Traditionally, there are three methods for adjusting covariates in survival data: matching, stratification and semi-parametric models (such as Cox model). Methods of matching and stratification have been discussed for many studies (Henkey, Myers, 1971; Cupples et al., 1995; Amato, 1988; Nieto, Coresh, 1996; Winnett, Sasieni, 2002). The basic idea for matching and stratification is to stratify the patients into strata according to the patients' confounding variables, so within each stratum, patients have similar characteristics. Then summarize the survival estimates for each stratum. However, there are some limitations for matching and stratification, it is often difficult to obtain well-matched data for studies with small sample sizes. Also, if there are continuous confounders or too many confounding variables, some strata can have too few individuals to analyze.

The Cox proportional hazards model is widely applied to survival data with a single endpoint. Recently, this approach has been extended to the competing risks setting. Wei, Lin and Weissfeld (1988) proposed a method of modeling marginal distributions using the Cox model for multivariate incomplete failure time data. Lunn and McNeil (1995) used the Cox regression model by accounting for the censoring in two different ways. One procedure runs Cox regression stratified by type of failure. The other uses the unstratified Cox regression, assuming that the hazard functions associated with the two types of failure have a constant ratio. Cheng, Fine and Wei (1998) introduced a method to construct confidence intervals and bands for the CIF under the Cox model. The approach was illustrated with data from a prostate cancer trial. All of these methods are based on modeling the cause specific hazard (also called "subdistribution hazard") using the Cox proportional hazards model. For modeling the cause, r, of cause specific hazard of an individual with covariate vector Z

$$\lambda^r(t|Z) = \lambda_0^r exp(\beta^{rT} Z),$$

where λ_0^r is the baseline cause specific hazard and β^r is the coefficient vector for cause r. Unlike survival data without competing risks, we can not construct the ratio of the cumulative incidence function based on Cox model estimates because the CIF is the function of hazard of all failure types.

Fine and Gray (1999) proposed a kind of hazard function, which can be attached to the cumulative incidence function. They defined their "subdistribution hazard" as

$$\begin{aligned}\lambda^1(t;Z) &= \lim_{\Delta t\to 0}\frac{1}{\Delta t}Pr\{t\le T\le t+\Delta t, \epsilon=1|T\ge t\cup(T\le t\cap\epsilon\ne 1), Z\}\\ &= -dlog\{1-I^1(t;Z)\}/dt,\end{aligned}$$

where $I^1(t;Z)$ is the cause 1 cumulative incidence function at time t with covariate Z. The risk set for the hazard λ^1 is unnatural, patients failed before time t due to other causes are still in the risk set. They also adapted the inverse probability of censoring weight (IPCW) technique to construct a partial likelihood function for right censored data.

There are two main disadvantages of regression models for adjusting confounders. First, these models focus on modeling the cause specific hazard, it is not adequate for describing the overall failure probability. Second, covariates can affect cause specific hazard and failure probability differently.

Recently, the "propensity score" technique is developed for adjusting confounders. It is defined as the conditional probability of being in the treatment, exposure or risk group given the observed covariates. The main idea is: replace the collection of confounding variables with a single value, which is a function of these covariates representing the "propensity" to receive the treatment. The propensity score can serve as a "balanced score" for studies with unbalanced covariates among groups. The inverse probability weighting approach was originally used in methods to determine the propensity score (Rosenbaum and Rubin, 1983, 1985). Later, inverse probability weighting was widely applied to many research areas, including the analysis of incomplete data (Wang et al., 1997), the estimation of causal effects (Dawson and Lavori, 2002; Lunceford and Davidian, 2004), and the analysis of survey data (Little, 1986).

Xie and Liu (2005) proposed an adjusted Kaplan-Meier estimator (AKME) using the inverse probability weight to adjust the unbalanced confounders. For the ith individual in the group k with the covariate Z_i, the weight is: $w_{ik}=\frac{1}{p_{ik}}=\frac{1}{P(X_i=k|Z_i)}$. After assigning the corresponding weight to each individual, they use the weighted number of events and weighted number at risk to obtain the adjusted Kaplan-Meier estimator. The following formula defines the AKME for the kth group:

$$\widehat{S}_k^w(t) = \begin{cases} 1 & \text{if } t < t_1 \\ \prod_{t_j \le t}[1 - d_{jk}^w / Y_{jk}^w] & \text{if } t_1 \le t \end{cases},$$

The adjusted Kaplan-Meier estimator outperforms the Kaplan-Meier estimator as well as other estimators based on stratification approaches. The strengths of the inverse probability weight are: it has good bias reduction and can take account of many covariates. Moreover, it can be easily performed.

3.6 SURVIVAL DATA ANALYSIS WITH DEPENDENT CENSORING

For survival data with dependent censoring, we can not ignore the information lost due to censoring. When dependent censoring is present, the usual estimates of survival can be biased and misleading. Much work has been done to address this issue. Roughly speaking, there are two types of approaches in this area depending on whether auxiliary prognostic factors are available or not. When the auxiliary variables are available, under the strong assumption that the censoring probability can be predicted by these time-dependent and independent auxiliary variables, the auxiliary variables are incorporated in order to recover information lost due to censoring and to improve survival estimates (Robins, 1993; Robins and Finkelstein, 2000; Schafstein and Robins, 2002; Rotnitzky et al.,2007). Some other useful estimation techniques are also proposed in various studies (Gray, 1992; Finkelstein and Schoenfeld, 1994; Fleming et al.,1994; Malani, 1995; Murray and Tsiatis, 1996). When the auxiliary variables are unavailable, based on various assumptions of the dependence structure of failure time and censoring time, by modeling the joint distribution of failure and censoring time, many authors evaluated the sensitivity of inference and generated bounds for the survival function (Slud and Rubinstein, 1983; Klein and Moeschberger, 1988, 1992; Zheng and Klein, 1994, 1995).

Robins and Finkelstein (2000) developed a weighted Kaplan-Meier estimator using the inverse probability of censoring weight (IPCW) (Robins, 1993). First, they constructed the IPCW KM estimator by incorporating the auxiliary variables. Then, they define the weight

with this estimator and calculate the weighted survival estimator in the absence of censoring. The interesting point is that they use the IPCW to construct the weight function and to incorporate the weight function into the usual Kaplan-Meier formula. Grunkemeier et al. (2007) proposed a similar method as Robins and Finkelstein's (2000). They considered the death from competing risks as dependent censoring and assigned the IPCW weight to each observed interest event, so that the patients censored due to competing risks were still at risk for the interest event. The IPCW weight they used was defined as an inverse of the estimated probability of each interest event being observed given all of the individual's characteristics and was derived from the Cox model. The big question in this approach is how to accurately estimate the probability of each interest event being observed. And accuracy of the IPCW estimation is strongly dependent on the sufficiency of the auxiliary variables. Specifically, if the auxiliary variables are sufficient, the hazard of the censoring of event 1 does not depend on the missing survival time due to the competing risks given the auxiliary variables.

4.0 PROPOSED STUDY I: ADJUSTED CUMULATIVE INCIDENCE FUNCTION

4.1 NOTATION

We begin by defining the notation. For an individual i ($i = 1, ..., N$), let T_i be the possibly right-censored event time; let δ_i be the censoring indicator, with $\delta_i = 0$ if T_i is censored and with $\delta_i = 1$ if T_i corresponds to an event. Let R_i indicate which competing events the individual experienced ($R_i = r$, where $r = 1, ..., M$ for M different competing events); let X_i indicate which group the individual belongs to ($X_i = 1, ..., K$ for K different groups); and let Z_i be the covariate vector. Therefore, $(T_i, \delta_i, R_i, X_i, Z_i)$ denotes an independent sample of possibly right-censored survival data with K groups.

The cumulative incidence function (CIF), which is the probability of failure from cause $R = r$ up to a certain time point t, will be denoted as $I^r(t) = P(T \leq t, R = r)$ with $r = 1, ..., M$. By using the conditional probability, the CIF can be rewritten in the form:

$$\begin{aligned} I^r(t_j) = P(T \leq t_j, R = r) &= \sum_{j'=1}^{j} P(T = t_{j'}, R = r) \\ &= \sum_{j'=1}^{j} P(T = t_{j'}, R = r | T \geq t_{j'}) P(T \geq t_{j'}) = \sum_{j'=1}^{j} \lambda^r(t_{j'}) S(t_{j'-1}). \end{aligned} \tag{4.1}$$

To obtain the CIF estimator, $\hat{I}^r(t)$, we substitute $S(t)$ in equation (1) with the Kaplan-Meier estimator $\hat{S}(t)$ and we substitute $\lambda^r(t_j)$ in equation (1) with the cause-specific hazard estimator $\hat{\lambda}^r(t_j) = d_j^r / Y_j$. The resulting estimated CIF has two properties. One is that the CIF is directly estimable without making any assumptions about the joint distribution of the potential failure time of the competing risks. The other is that $F_i(t)$ is not a true

distribution function, since $F_i(\infty) = P(\delta = i)$. It has a property that is nondecreasing with $F_i(0) = 0$ and $F_i(\infty) < 1$. Such a function is also called a "subdistribution" function.

4.2 PROPOSED FUNCTION

To obtain the weight, we begin by determining the probability that each individual participant will be in a particular group, conditional on the individual's characteristics. Individuals who have a high probability of being in a particular group will be treated as overrepresenting the group and will therefore receive a lower weight. Individuals who have a low probability of being in a particular group will be treated as under representing the group and will therefore receive a higher weight. The weight for any individual for a given group is therefore equal to the inverse of the probability of being in the group. The weight then is used to adjust the cause-specific hazard estimator and the Kaplan-Meier survival estimator. Our proposed ACIF estimator is formed using these two adjusted estimators.

To construct a weight for each individual, let p_{ik} be the probability of the ith individual being in group k. This probability may depend on the covariate vector Z_i, i.e., $p_{ik} = P(X_i = k \mid Z_i)$. We will assign the ith individual a weight of $w_{ik} = 1/p_{ik}$ when this individual is in the kth group. If there are only two groups involved (such as treatment $k = 1$ and control $k = 0$), then the weight for individual i will be assigned as

$$\begin{cases} w_{i1} = \frac{1}{p_{i1}} = \frac{1}{P(X_i=1|Z_i)} & \text{if patient } i \text{ is in the treatment group; and} \\ w_{i0} = \frac{1}{p_{i0}} = \frac{1}{P(X_i=0|Z_i)} & \text{if patient } i \text{ is in the control group.} \end{cases} \tag{4.2}$$

After assigning a weight to each individual, we use the following three steps to construct the proposed ACIF:

1. Find the adjusted overall survival function.
2. Find the adjusted cause-specific hazard function.
3. Form the ACIF by using the adjusted overall survival function and cause-specific hazard function from the previous two steps.

Suppose in a sample that there are D distinct event times $t_1 < t_2 < ... < t_D$. At time t_j $(j = 1, ..., D)$, in group k there are d_{jk} events among Y_{jk} individuals who are at risk. Note that the risk set Y_{jk} is defined as the number of those who have not experienced any event by time t_j. For each time t_j, we let d^w_{jk} be the weighted number of events and Y^w_{jk} be the weighted number of individuals at risk at time t_j in group k.

The overall survival function $S(t)$ is defined as the probability of being "event-free" at time t, where an "event" means any event (the event of interest or a competing event). If there is only one end point (no competing events), the estimated overall survival function can be obtained by the Kaplan-Meier estimator. When there is a confounding variable involved, the Kaplan-Meier estimator needs to be adjusted to avoid bias. Xie and Liu (2005) proposed an adjusted overall survival function for this situation. To adapt their idea to the case of multiple end points, we propose to estimate the overall survival function for group k by the form:

$$\widehat{S}^w_k(t) = I(t < t_1) + I(t \geq t_1) \cdot \prod_{t_j \leq t} [1 - d^w_{jk}/Y^w_{jk}],$$

where the weighted number of events are

$$d^w_{jk} = \sum_{i:T_i = t_j} w_{ik}\delta_i I(X_i = k) = \sum_{i:T_i = t_j} \frac{\delta_i I(X_i = k)}{p_{ik}}$$

and the adjusted risk sets are

$$Y^w_{jk} = \sum_{i:T_i \geq t_j} w_{ik} I(X_i = k) = \sum_{i:T_i \geq t_j} \frac{I(X_i = k)}{p_{ik}}.$$

To obtain the adjusted cause-specific hazard function, let r be the event of interest $(r \in \{1, ..., M\})$, let d^{rw}_{jk} denote the weighted number of events of interest, and again let Y^w_{jk} be the weighted number at risk for group k at time t_j. We define the adjusted cause-specific hazard for the kth group at time t_j to be $\hat{\lambda}^{rw}_k(t_j) = d^{rw}_{jk}/Y^w_{jk}$, where the weighted number of events of interest has the form

$$d^{rw}_{jk} = \sum_{i:T_i = t_j} w_{ik}\delta_i I(\varepsilon_i = r) I(X_i = k) = \sum_{i:T_i = t_j} \frac{\delta_i I(\varepsilon_i = r) I(X_i = k)}{p_{ik}}.$$

Finally, we use the overall survival function at time t_{j-1} and the cause-specific hazard at time t_j to give the form of the ACIF by summing up the product of the overall survival and cause-specific hazard over time $t_1, t_2, ..., t_j$. This is the estimator for the probability of cause r for group k after the unbalanced covariates are taken into account. This process can be written as:

$$\hat{I}_k^{rw}(t_j) = \sum_{j'=1}^{j} \hat{\lambda}_k^{rw}(t_{j'}) \cdot \hat{S}_k^{w}(t_{j'-1}). \tag{4.3}$$

Note that the proposed ACIF in equation (2) tries to add weight to each observation in the study, where the weight is the inverse probability of the individual being assigned to group k. When each individual has the same weight (meaning that each individual has an equal probability of being in each group), then we can prove that the ACIF is reduced to the CIF estimator.

4.3 PROPERTIES

In Appendix A, we show that given information up to time t_j, $E_j\{\hat{I}_k^{rw}(t_j)\} = I_k^{rw}(t_j)$ where E_j denotes a conditional expectation. Therefore, in the range where we have data, our estimated ACIF is an unbiased estimator for the underlying ACIF.

The estimated variance of the CIF has been discussed in several articles (Dinse and Larson, 1986; Gaynor et al., 1993; Klein and Moeschberger, 2003). Among them, the formula from Gaynor et al. (1993) is more intuitive. In this article, we use a similar expression for the estimated variance of our ACIF. The estimated variance of $\hat{I}^{rw}(t)$, $\widehat{Var}\{\hat{I}^{rw}(t)\}$ is equal to

$$\sum_{i=1}^{j} \widehat{Var}\left\{\hat{\lambda}_i^{rw}\hat{S}^{w}(t_{i-1})\right\} + 2\sum_{i=1}^{j-1}\sum_{i'=i+1}^{j} \widehat{Cov}\left\{\hat{\lambda}_i^{rw}\hat{S}^{w}(t_{i-1}),\ \hat{\lambda}_{i'}^{rw}\hat{S}^{w}(t_{i'-1})\right\}.$$

In Appendix B, we show that the variance and covariance in the above equation can be estimated by

$$\widehat{Var}\left\{\hat{\lambda}_i^{rw}\hat{S}^{w}(t_{i-1})\right\} = \left\{\hat{S}^{w}(t_{i-1})\right\}^2 (\hat{\lambda}_i^{rw})^2 \left[\frac{1-\hat{\lambda}_i^{rw}}{M_i\hat{\lambda}_i^{rw}} + \sum_{l=1}^{i-1}\left\{\frac{1-\hat{s}^{w}(t_l)}{M_l\hat{s}^{w}(t_l)}\right\}\left(\frac{1-\hat{\lambda}_i^{rw}}{M_i\hat{\lambda}_i^{rw}}+1\right)\right],$$

and

$$\begin{aligned}
&\widehat{Cov}\left\{\hat{\lambda}_i^{rw}\hat{S}^w(t_{i-1}),\ \hat{\lambda}_{i'}^{rw}\hat{S}^w(t_{i'-1})\right\} \\
= \ & \hat{\lambda}_i^{rw}\hat{\lambda}_{i'}^{rw}\hat{S}^w(t_{i-1})\hat{S}^w(t_{i'-1})[\{\sum_{l=1}^{i-1}\frac{1-\hat{s}^w(t_l)}{M_l\hat{s}^w(t_l)}+1\}\{1-\frac{1-\hat{\lambda}_i^{rw}}{M_i\hat{s}^w(t_i)}\}-1],
\end{aligned}$$

respectively, where $M_i = (\sum_{j:T_j\geq t_i} 1/p_j)^2 / \sum_{j:T_j\geq t_i}(1/p_j)^2$ and $s^w(t_l) = S^w(t_l)/S^w(t_{l-1})$. If each person in the study has equal weight, we can have $M_i = Y_i$, and the variance and covariance in the above equation can be further simplified as

$$\widehat{Var}\left\{\hat{\lambda}_i^r\hat{S}(t_{i-1})\right\} = \left\{\hat{S}(t_{i-1})\right\}^2(\hat{\lambda}_i^r)^2\left[\frac{Y_i-d_i^r}{Y_id_i^r}+\left\{\sum_{l=1}^{i-1}\frac{d_l}{Y_l(Y_l-d_l)}\right\}\left(\frac{Y_i-d_i^r}{Y_id_i^r}+1\right)\right],$$

and

$$\begin{aligned}
&\widehat{Cov}\left\{\hat{\lambda}_i^r\hat{S}(t_{i-1}),\ \hat{\lambda}_{i'}^r\hat{S}(t_{i'-1})\right\} \\
= \ & \hat{\lambda}_i^r\hat{\lambda}_{i'}^r\hat{S}(t_{i-1})\hat{S}(t_{i'-1})\left[\sum_{l=1}^{i-1}\frac{d_l}{Y_l(Y_l-d_l)}-\frac{Y_i-d_i^r}{Y_i(Y_i-d_i)}\left\{1+\sum_{l=1}^{i-1}\frac{d_l}{Y_l(Y_l-d_l)}\right\}\right].
\end{aligned}$$

This formula can be used to estimate the variance of CIF, which is very similar to Gaynor's (Gaynor et al., 1993). The $(1-\alpha)\times 100\%$ pointwise confidence interval for an ACIF is $\hat{I}^{rw}(t) \pm Z_{1-\alpha/2}\cdot\ \widehat{Var}[\hat{I}^{rw}(t)]^{1/2}$.

4.4 INFERENCE TEST

Consider the following hypotheses for testing the equality of two ACIFs:

$$H_0\text{: } I_1^{rw}(t) = I_2^{rw}(t) \text{ for all } t \leq \tau \text{ and}$$
$$H_a\text{: } I_1^{rw}(t) \neq I_2^{rw}(t) \text{ for some } t \leq \tau,$$

where τ is the longest time at which both groups have at lease one participant at risk.

The test statistic can be written as

$$WI = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \cdot \int_0^\tau K(t) \left\{ \hat{I}_1^{rw}(t) - \hat{I}_2^{rw}(t) \right\} dt, \tag{4.4}$$

where $K(t)$ is a suitably chosen weight function. This weight function should be a function of participants at risk in each group at time t, and it should give a lower weight to the later time points because of the decreasing weight with increasing time. By using $K(t)$, it is easy to detect an early difference and to stabilize the test statistic.

Pepe (1991) showed that under the null hypothesis of equality, the test statistic is asymptotically normal and has a mean of 0. Thus, the test statistic $Z_0 = WI/\sqrt{Var(WI)}$ has a standard normal distribution for large samples under the null hypothesis. Specifically, we reject the null hypothesis at α level when $|Z_0| > Z_{\alpha/2}$, where $Z_{\alpha/2}$ is the critical value of the standard normal distribution.

The hypothesis test of CIFs has been discussed by many researchers (Gray, 1988; Pepe, 1991; Pepe and Mori, 1993; Lin, 1997). Lin (1997) proposed a resampling technique for developing the approximate distribution of the CIF estimator. Based on the approximate distribution, Lin built the confidence bands for CIF and constructed Kolmogorov-Smirnov type tests for comparing two CIF curves. For data with a small or moderate sample size, we developed a bootstrap approach to test our hypotheses. Given known data with n observations, we observe individual i as (T_i, d_i, X_i, Z_i), where T_i is the survival time; d_i is the event index; X_i is the group index; and Z_i is the covariate of the ith participant. The steps of the bootstrap approach are as follows:

1. From the n observations, (T_1, d_1, X_1, Z_1), (T_2, d_2, X_2, Z_2), $\cdots$, (T_n, d_n, X_n, Z_n), we sample n times with replacement from this data and form a single bootstrap data set, $(T_{1^*}, d_{1^*}, X_{1^*}, Z_{1^*})$, $(T_{2^*}, d_{2^*}, X_{2^*}, Z_{2^*})$, $\cdots$, $(T_{n^*}, d_{n^*}, X_{n^*}, Z_{n^*})$. For each bootstrap sample, we use logistic regression to calculate the corresponding weight, w_{1^*}, w_{2^*}, $\cdots$, w_{n^*}.
2. We use the single bootstrap data obtained from step 1 to calculate the estimated ACIF functions for the event of interest and for the treatment and control groups, $\hat{I}_1^{1w}(t)$ and $\hat{I}_2^{1w}(t)$, at each time point. We calculate the test statistic $WI^{(B)} = \sqrt{n_1 n_2/(n_1+n_2)}$ $\sum K(t)\{\hat{I}_1^{1w}(t) - \hat{I}_2^{1w}(t)\}$.
3. We repeat steps 1 and 2 for B times.
4. From the B bootstrap samples, we calculate the variance of $WI^{(B)}$, $\widehat{Var}\{WI^{(B)}\}$. We let $Z_B = WI^{(B)}/\sqrt{\widehat{Var}\{WI^{(B)}\}}$. Thus, the two-sided P-value of the hypothesis test is $2 \cdot Pr(|Z_B| > Z_{\alpha/2}) = (1/B) \cdot \sum_B I(|Z_B| \geq |Z_{\alpha/2}|)$.

4.5 SIMULATION STUDIES

4.5.1 Performance of ACIF estimators

In this section, we assessed the performance of the ACIFs for scenarios with highly unbalanced, intermediate unbalanced, and balanced covariate groups. A sample size of 400 was used. First, we generated the group indicator variable X_i with $X = 1$ and $X = 0$ 200 times each. Second, we generated the covariate Z_i for individual i ($i = 1, 2, \cdots, 400$) from a Bernoulli distribution with probabilities $P = (X = 1|Z = 1) = 0.8$ and $P = (X = 1|Z = 0) = 0.2$ for highly unbalanced data; with probabilities $P = (X = 1|Z = 1) = 0.6$ and $P = (X = 1|Z = 0) = 0.4$ for intermediate unbalanced data; and with probabilities $P = (X = 1|Z = 1) = 0.5$ and $P = (X = 1|Z = 0) = 0.5$ for balanced data. Third, let T_1 and T_2 denote two lognormally distributed random variables, where $T_j = e^{Y_j}$ and (Y_1, Y_2) follows a bivariate normal distribution. Let u_j and σ_j denote the mean and standard deviation of Y_j. We transformed the correlation ρ between Y_1 and Y_2 into the Spearman correlation $r = (6/\pi)arcsin(\rho/2)$, such that this number was both the Spearman correlation

between Y_1 and Y_2 as well as between T_1 and T_2. In our simulation, we used $\rho = 0.5$ for the correlation between Y_1 and Y_2. If $Z = 1$, $u_1 = 2.5$, $u_2 = 1$, $\sigma_1 = \sigma_2 = 1$; if $Z = 0$, $u_1 = 0.5$, $u_2 = 1$, $\sigma_1 = \sigma_2 = 1$. Fourth, create the censoring sample from a uniform distribution. For the three scenarios, we investigated three censoring rates: 0, 0.2 and 0.4.

To calculate the naive CIF, we used the CIF formula without considering the unbalanced covariate Z. To calculate the ACIF, we used equation (4.3). To calculate the true underlying CIF, we used all 400 samples without considering the group factor X because both groups should have equal CIF.

For each scenario, we plotted the naive CIF, the ACIF, and the true CIF for each covariate group (Figures 1-3). We also calculated the mean square difference between the naive CIF and the true CIF and the mean square difference between the ACIF and the true CIF (Table 1) over all failure times for the event of interest. As shown in Figures 1-3 and Table 1, the ACIF provided better estimates for the true underlying CIF than the naive CIF did. The groups that were more unbalanced demonstrated a larger bias when the naive CIF measures were used. When the percentage of censoring increased, this phenomenon became more pronounced. When the groups had balanced covariates, both the naive CIF and the ACIF gave good estimates of the underlying CIF.

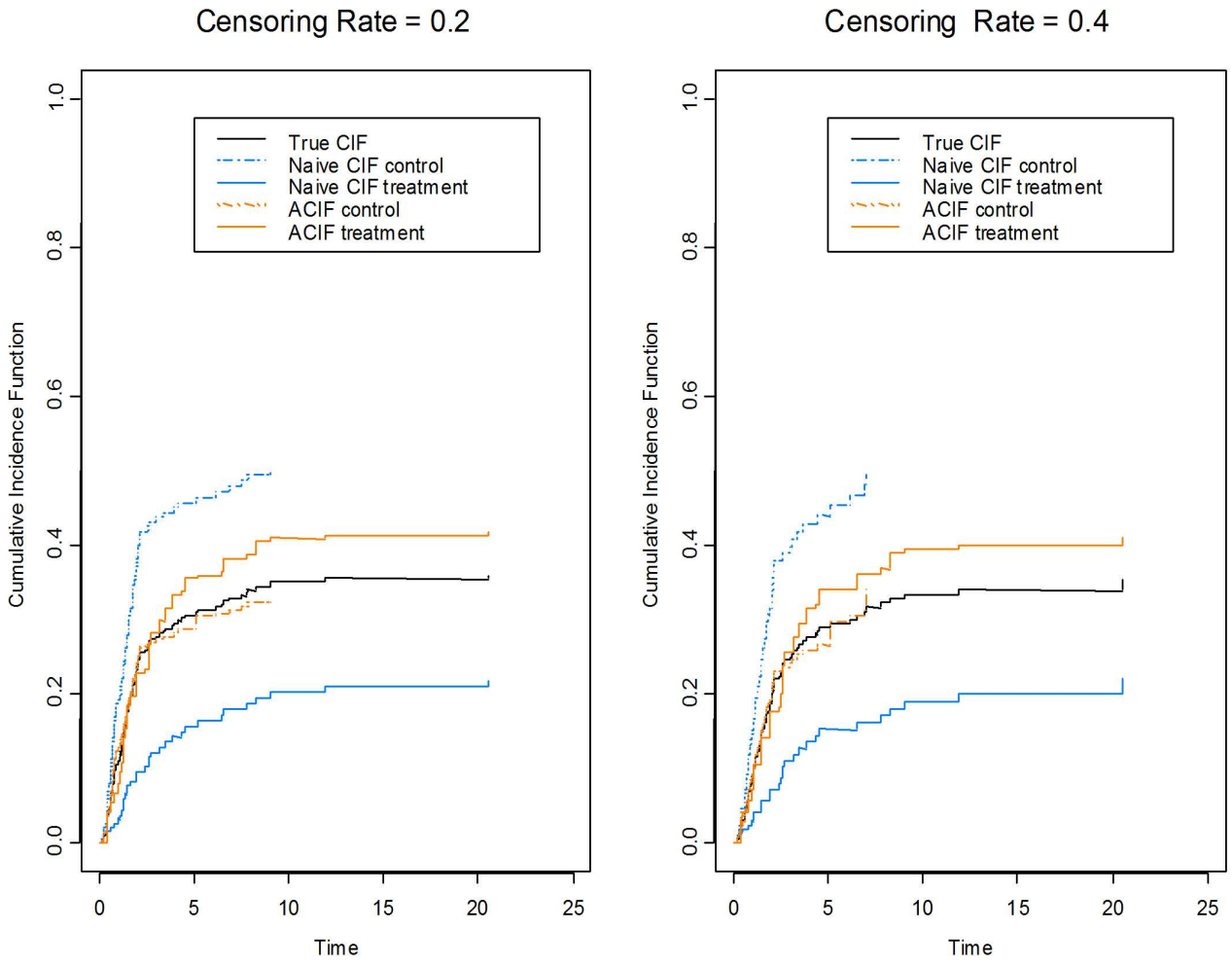

Figure 1. Comparison of the estimated naive CIF, the estimated ACIF, and the underlying true CIF with highly unbalanced covariate groups.

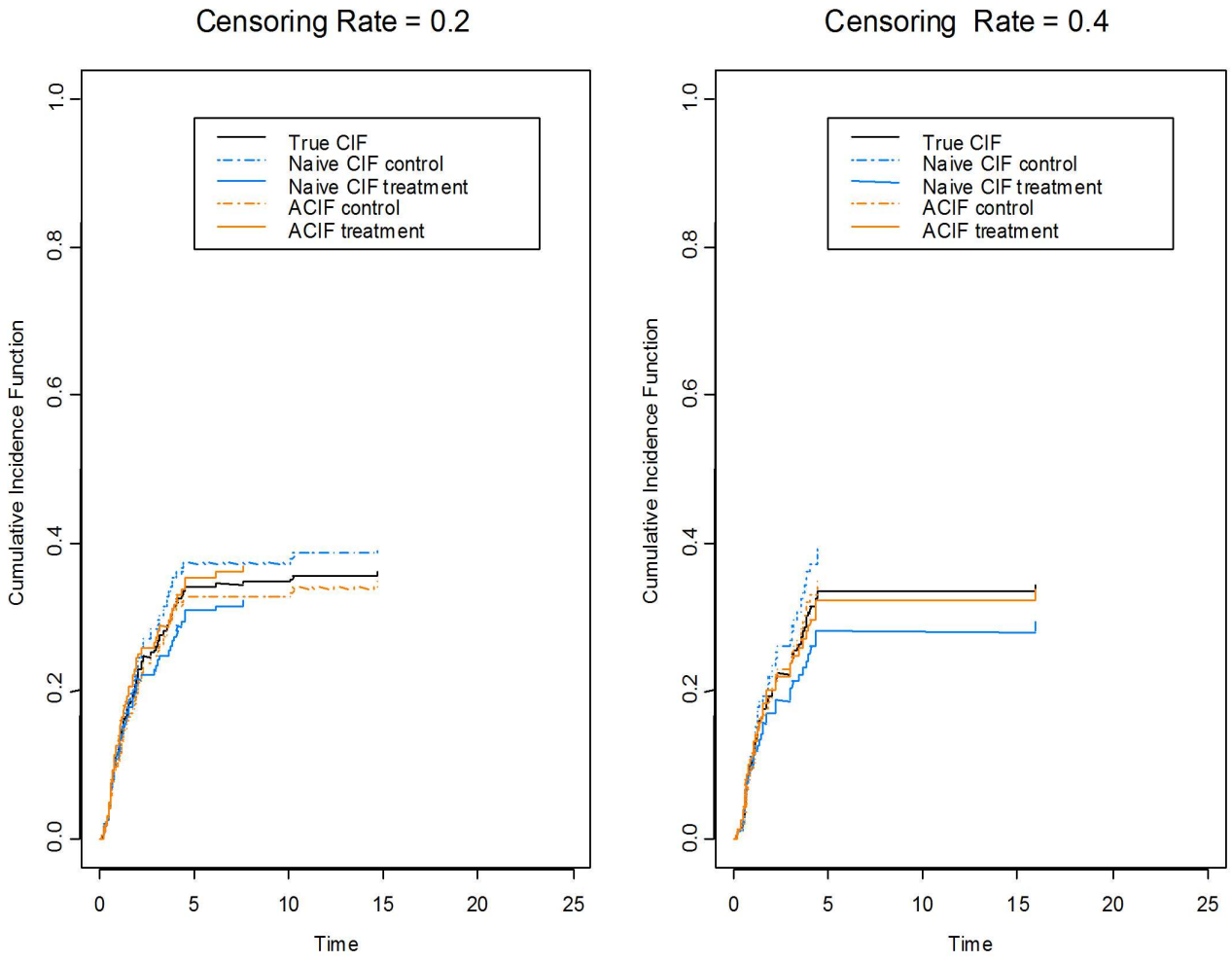

Figure 2. Comparison of the estimated naive CIF, the estimated ACIF, and the underlying true CIF with intermediate unbalanced covariate groups.

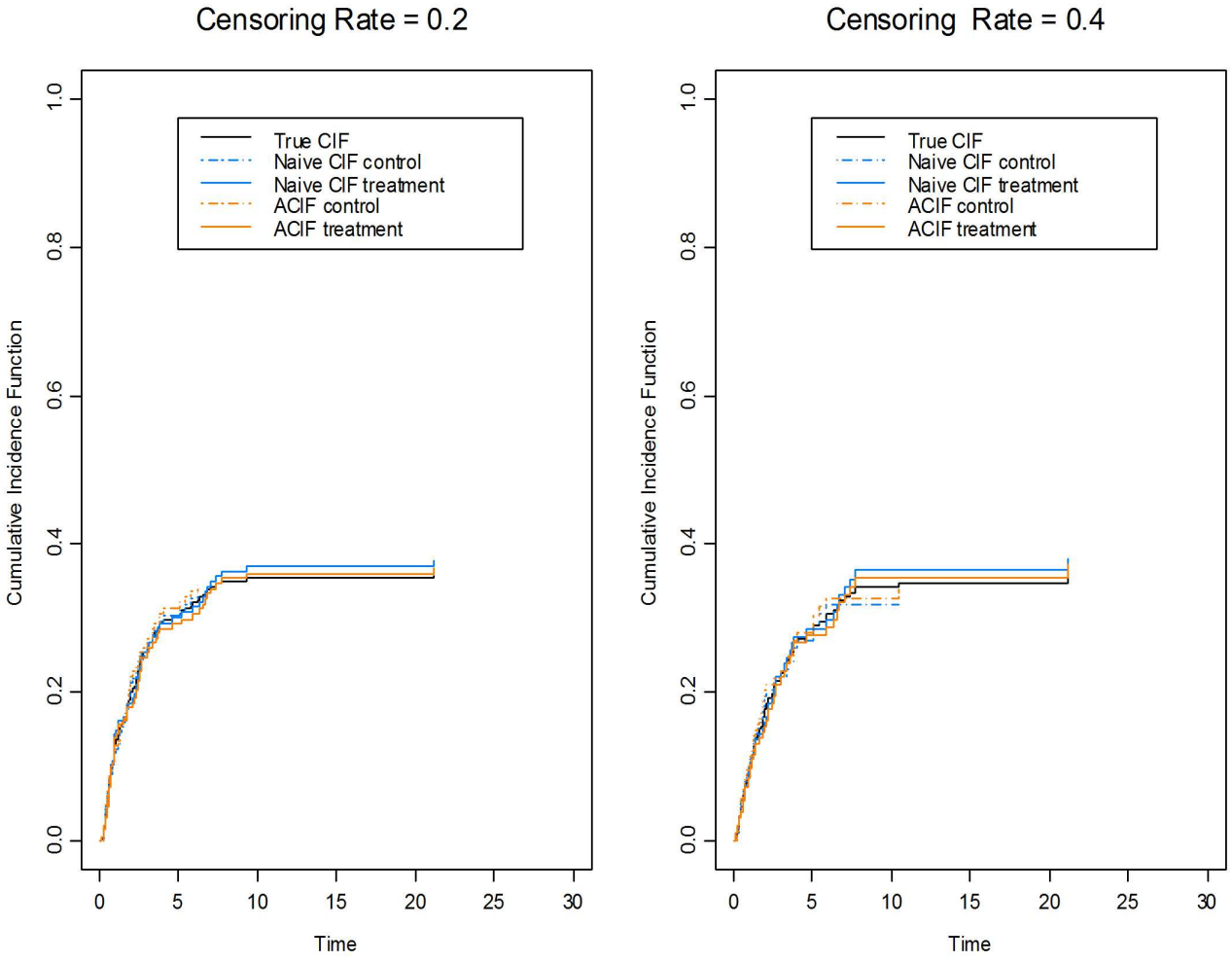

Figure 3. Comparison of the estimated naive CIF, the estimated ACIF, and the underlying true CIF with balanced covariate groups.

Table 1. Summary of the mean square difference for the estimated disease-specific cumulative incidence function (CIF) and the estimated disease-specific adjusted cumulative incidence function (ACIF)

		CIF		ACIF	
Distribution of covariate	Censoring rate	Treatment (X=1)	Control (X=0)	Treatment (X=1)	Control (X=0)
Highly unbalanced	0	0.014	0.011	0.00078	0.00004
	0.2	0.015	0.011	0.0011	0.00010
	0.4	0.014	0.011	0.0012	0.00005
Intermediate unbalanced	0	0.00033	0.00040	0.00015	0.00013
	0.2	0.00032	0.00042	0.00024	0.00013
	0.4	0.00071	0.00091	0.00015	0.00009
Balanced	0	0.00005	0.00004	0.00006	0.00005
	0.2	0.00005	0.00005	0.00007	0.00012
	0.4	0.00009	0.00008	0.00010	0.00017

4.5.2 Performance of Variance Estimator of ACIF

In this section, we assessed the performance of the variance estimators of ACIFs for scenarios with highly unbalanced, intermediate unbalanced, and balanced covariate groups. The simulation data were generated using the same procedures as described in section 4.5.1. A sample size of 400 and censoring rates of 0 and 0.2 were used. Independent samples were generated for 500 iterations and three time points were investigated, $t_0 = 1.0, 3.0$ and 5.0.

The variance estimators at each time point were calculated using our formula from 500 independent iterations. The mean of these variance estimators was compared with the empirical variance of ACIF (Table 2, Table 3). For both censoring rates, our estimators were very close to the empirical variance for balanced covariate groups. For the unbalanced covariate groups, our estimators seemed to perform well while the empirical variances were poorly estimated due to the dependent variance among groups. Moreover, the more unbalanced covariate was, the larger discrepancy existed between the estimators and the empirical variances. The empirical variances we obtained were calculated from the variance of the ACIF for each group from 500 independent iterations. While within each iteration, we calculated the weight for each individual and split this dataset into two subsets for each of two groups. Thus, the two subsets are dependent due to the weighting procedure and the ACIF for each of the two groups are dependent to each other. Eventually, the empirical variance for two groups were dependent due to the unbalanced covariates distribution among groups. Therefore, the traditional method of empirical variance was not valid for estimating the variance of ACIF. In contrast, our variance estimator could be consistent for practical applications.

Table 2. Summary of the performance of variance estimator compared with the empirical variance (censoring rate = 0)

Distribution of covariate	Time point t_0	Control (X=0) $\hat{I}_1^{1w}$	var $(\hat{I}_1^{1w})$	mean $(\widehat{var}(\hat{I}_1^{1w}))$	Treatment (X=1) $\hat{I}_0^{1w}$	var $(\hat{I}_0^{1w})$	mean $(\widehat{var}(\hat{I}_0^{1w}))$
Balanced	1.0	0.137	0.000538	0.000583	0.138	0.000561	0.000590
	3.0	0.293	0.000825	0.000971	0.293	0.000928	0.000970
	5.0	0.339	0.000869	0.000994	0.338	0.000997	0.000992
Intermediate unbalanced	1.0	0.138	0.000452	0.000624	0.139	0.000697	0.000609
	3.0	0.295	0.000670	0.00105	0.294	0.000980	0.000976
	5.0	0.341	0.000817	0.00109	0.339	0.000987	0.000990
Highly unbalanced	1.0	0.137	0.000362	0.000964	0.139	0.00140	0.000887
	3.0	0.295	0.000574	0.00170	0.294	0.00159	0.00136
	5.0	0.340	0.000713	0.00178	0.339	0.00155	0.00135

Table 3. Summary of the performance of variance estimator compared with the empirical variance (censoring rate = 0.2)

Distribution of covariate	Time point t_0	Control (X=0) $\hat{I}_1^{1w}$	Control (X=0) var $(\hat{I}_1^{1w})$	Control (X=0) mean $(\widehat{var}(\hat{I}_1^{1w}))$	Treatment (X=1) $\hat{I}_0^{1w}$	Treatment (X=1) var $(\hat{I}_0^{1w})$	Treatment (X=1) mean $(\widehat{var}(\hat{I}_0^{1w}))$
Balanced	1.0	0.119	0.000589	0.000571	0.120	0.000501	0.000572
	3.0	0.270	0.00101	0.00109	0.271	0.000786	0.00110
	5.0	0.318	0.00119	0.00118	0.319	0.000955	0.00118
Intermediate unbalanced	1.0	0.121	0.000458	0.000584	0.119	0.000646	0.000582
	3.0	0.272	0.000820	0.00114	0.271	0.00121	0.00109
	5.0	0.320	0.000998	0.00116	0.319	0.00125	0.00116
Highly unbalanced	1.0	0.119	0.000343	0.000939	0.121	0.00120	0.000856
	3.0	0.269	0.000747	0.00192	0.274	0.00189	0.00151
	5.0	0.318	0.00103	0.00213	0.322	0.00202	0.00156

4.6 ANALYSIS OF LIVER TRANSPLANT DATA

In this section, we use data derived from the Organ Procurement and Transplantation Network (2004) to demonstrate how an unbalanced distribution of a confounding variable affects the estimated CIF and to show how our proposed ACIF is used. The data are for candidates on the waiting list for a liver transplant. The candidates include 6,114 adults (aged 16 years or older) who had various types of end-stage liver disease, joined the waiting list at any time in year 2002, and were followed until January 31, 2004.

Our main covariate of interest was the variable indicating 10 disease types, and the confounding variable was the score derived from the model for end-stage liver disease (MELD score). In the analytic data set used here, we excluded 462 patients with unknown disease type, unknown MELD score, or unknown reasons for removal from the waiting list. Among the remaining 5,652 patients, 563 died before receiving a transplant (pre-transplant death), 2,308 received a transplant, 207 were removed from the waiting list for reasons other than death or transplant (e.g., they experienced an improvement in health that changed their need or desire for a transplant), and 2,574 were alive and still waiting for a transplant at the time of the study cutoff (Figure 4).

The distribution of the MELD scores for each of the 10 disease types is summarized in Table 4. Note that the MELD score ranges from 6 to 40, with 6 indicating the least ill patient and with 40 indicating the sickest patient. We found that patients with alcoholic liver disease and patients with metabolic liver disease had the highest mean MELD scores, while patients with liver cancer had the lowest mean MELD score. Patients with other liver disease types had similar mean MELD scores. For each of the 10 disease types, Figure 5 depicts the estimated CIF and Figure 6 depicts the estimated ACIF. As shown in Figure 7, although the estimated CIFs differed for patients with primary biliary cirrhosis and differed from those of patients with alcoholic liver disease ($P = 0.016$), the estimated ACIFs for these groups did not differ ($P = 0.23$). And although the estimated CIFs for patients with primary sclerosing cholangitis differed from those of patients with alcoholic liver disease ($P = 0.006$), the estimated ACIFs for these groups did not differ ($P = 0.24$).

To estimate the ACIFs and to test whether the CIFs or ACIFs were equal among the 10 disease groups, we expanded equation (4.2) and used multinomial logistic regression to obtain the weight for each patient by estimating the probability that the patient would have a specific type of liver disease given his or her MELD score. For the post hoc comparison of CIFs and ACIFs in paired groups, we used the weights obtained from regular binomial logistic regression. The bootstrap sample size B used in the example was 1,000.

Because the data might violate the proportional hazards assumption, we used the weight function $K(t) = nC_1^*(t)C_2^*(t)/\{n_1C_1^*(t) + n_2C_2^*(t)\}$ in equation (4.4) for the adjustment. This weight function is 1 minus the left continuous Kaplan-Meier estimator of the censoring distribution function, where $C_1^*(t)$ and $C_2^*(t)$ are the censoring distribution and where n_1 and n_2 are the sample sizes for the two disease groups.

Table 4. Scores Derived from the Model for End-stage Liver Disease (MELD Scores) and Distributed among Patients with Ten Types of Liver Disease

	MELD Score			
Liver Disease Type	*N*	Mean	SD	Median
Primary biliary cirrhosis (PBS)	253	14.70	7.04	13
Primary sclerosing cholangitis (PSC)	270	14.63	7.15	13
Hepatitis C and similar infections	2,127	14.90	6.73	13
Alcoholic liver disease	846	17.19	7.42	16
Hepatitis B	202	17.12	8.55	15
Autoimmune disorder	752	15.91	7.24	14
Liver cancer	292	13.02	6.11	12
Metabolic liver disease	105	18.31	8.89	17
Acute hepatic failure (AHF)	327	15.93	8.33	13
Other liver diseases	478	16.05	7.97	14

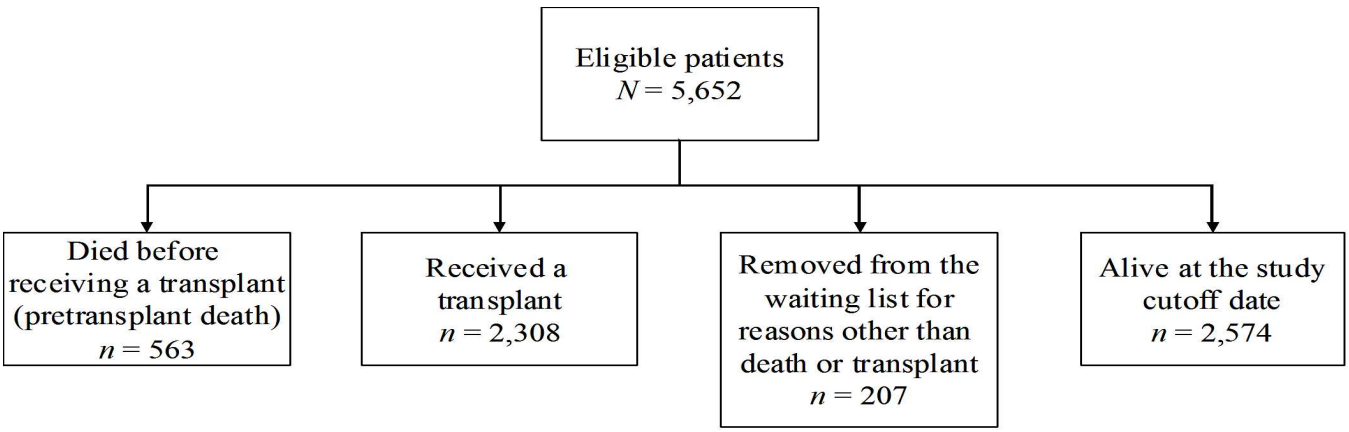

Figure 4. Conditions for the last follow-up.

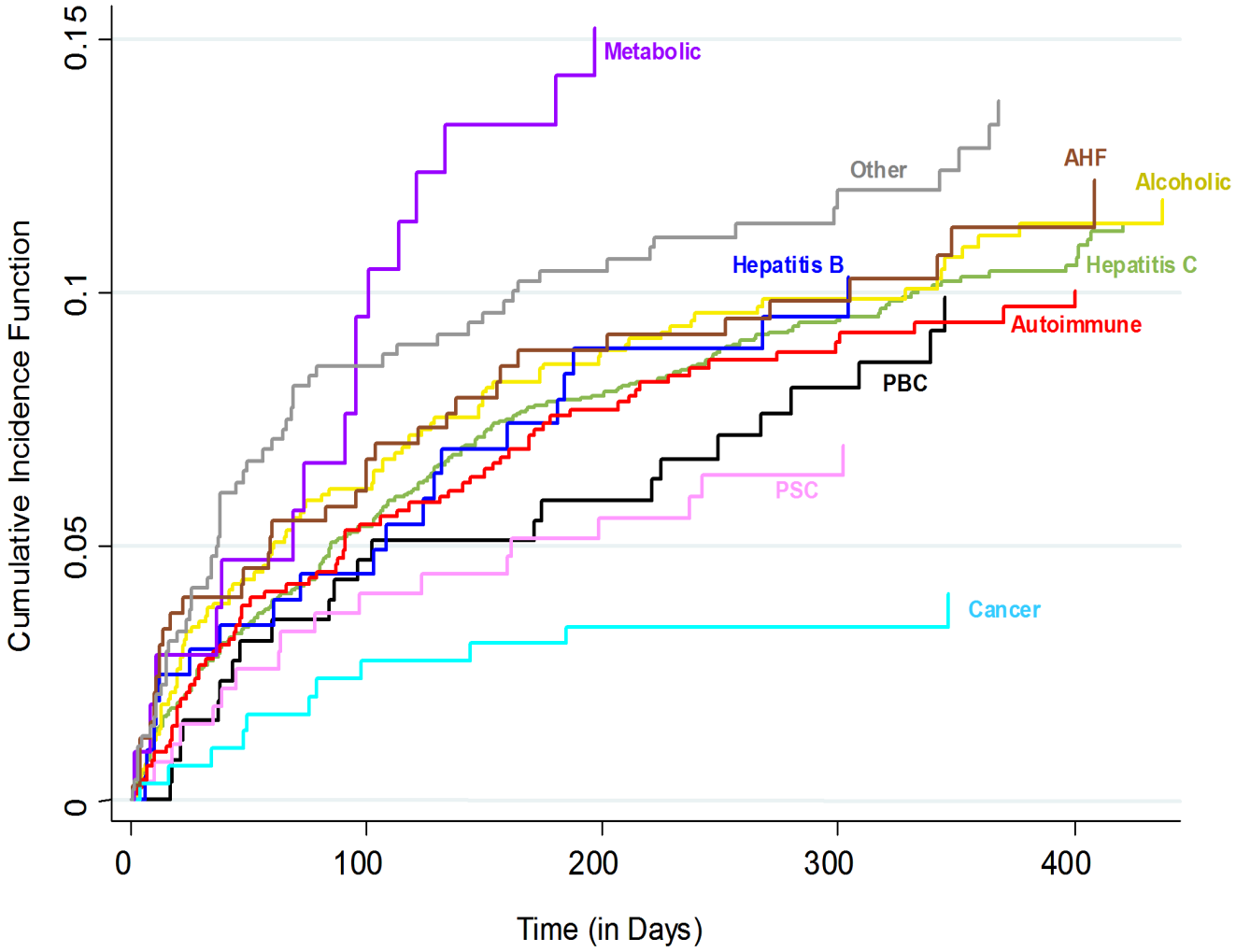

Figure 5. Estimated disease-specific cumulative incidence function.

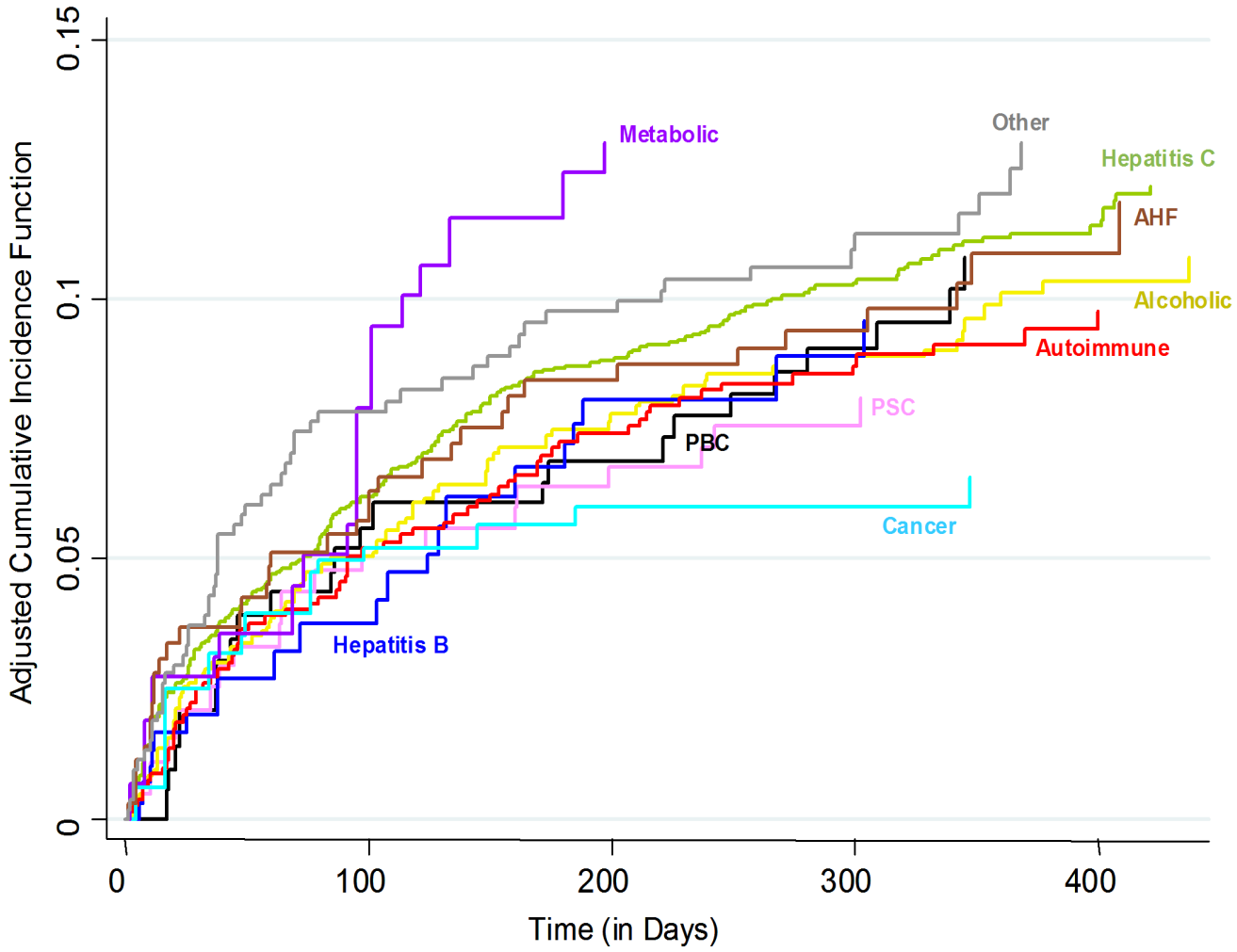

Figure 6. Estimated disease-specific adjusted cumulative incidence function.

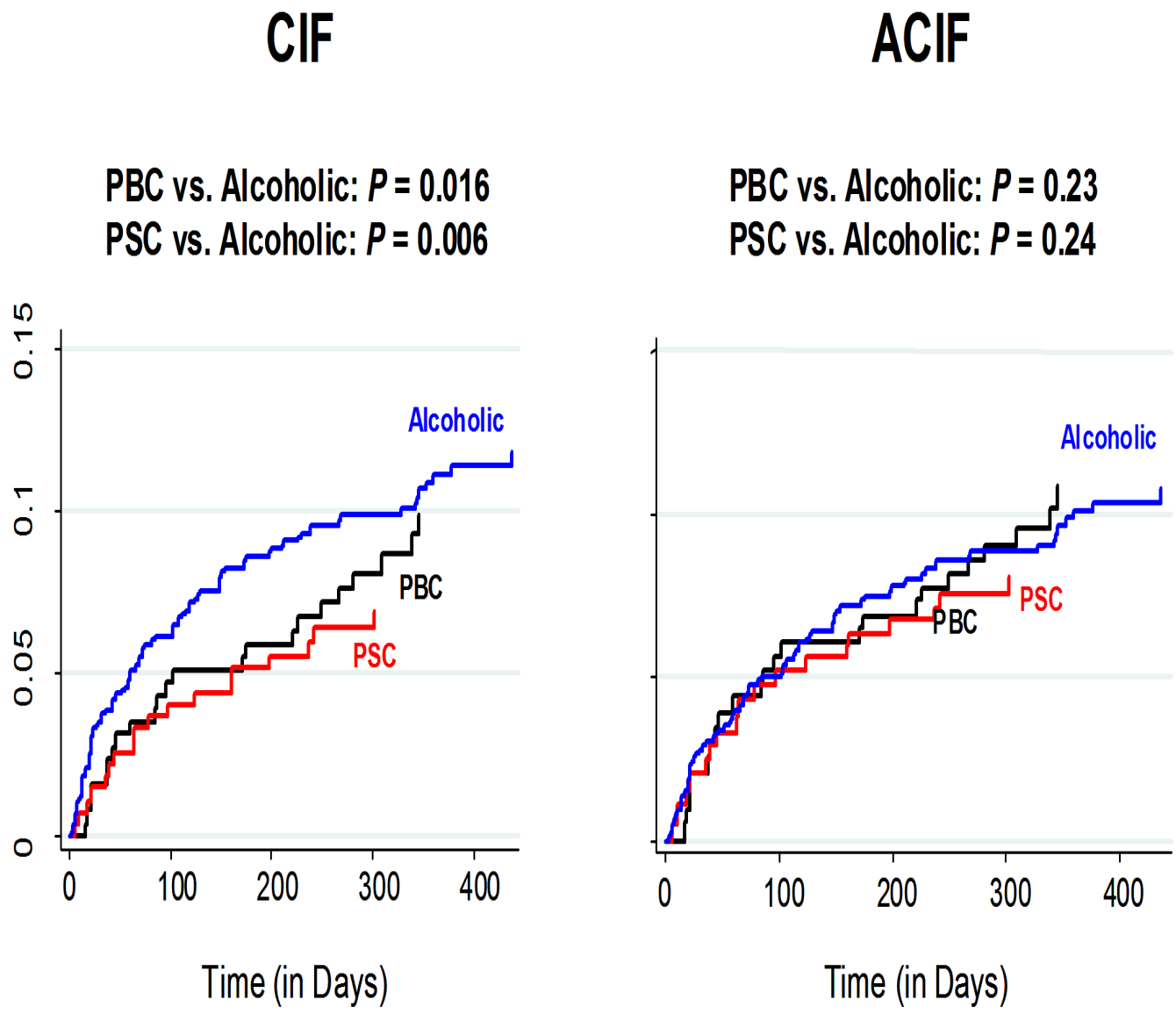

Figure 7. Comparison of the estimated disease-specific cumulative incidence function (CIF) and the estimated disease-specific adjusted cumulative incidence function (ACIF).

4.7 DISCUSSION

There are several different versions of the estimated variance of CIF. In general, these estimates can be classified as: the approaches based on counting process or martingale theory (Aalen, 1978; Pepe, 1991; Gray, 1988) and the approaches based on moments of the multinomial distribution (Dinse and Larson, 1986; Gaynor et al., 1993). Klein and Moeschberger (2003) proposed another kind of formula that estimated the variance as a function of the CIF. Braun and Yuan (2007) found that the multinomial-moment-based estimators outperformed than the others, especially for small sample size. Therefore, in our study, we adapted the idea of multinomial-moment-based approaches from Gaynor et al. (1993). When there are no unbalanced covariates involved among groups, our variance estimator can be simplified to a formula similar to Gaynor's. Our simulation study also showed that the traditional empirical variance from bootstrap methods performed poorly due to the dependence of ACIF among groups.

To compare two ACIFs, we developed a test statistic that is based on a generalized test for stochastic ordering of functions. This kind of test was originally proposed by Pepe and Fleming (1989), who used the non-parametric weighted Kaplan-Meier estimator to test for the difference between two survival functions. We are currently developing another test statistic that will be based on generalized linear rank statistics and will be a modification of Gray′s (1988) tests. For a discussion of the differences between statistics based on stochastic ordering and statistics based on linear rank tests, see Pepe and Fleming (1989, 1991).

In this study, we used only one example of the weight function $K(t)$ in equation (4.4). The weight function was a positive function, with lower weights reflecting smaller sample sizes as time progresses. Because the choice of $K(t)$ will affect the stability and interpretability of the results, we will be exploring other choices in future studies.

5.0 PROPOSED STUDY II: ADJUSTED SURVIVAL FUNCTION

5.1 INTRODUCTION

The cumulative incidence function is a marginal distribution, which estimates the cause specific failure probability when the patients are exposed to two or more risks simultaneously. Therefore, the CIF of the event of interest is affected by the occurrence of the competing event and it only presents partial information if the failure probabilities of the competing events do not present simultaneously.

For survival data with competing risks, most of the time we still want to answer: "what is the failure probability of our cause of interest if there are no competing risks?". In this case, the net survival probability $S^1(t) = Pr(T_1 \geq t)$ is more appropriate and it is a marginal survival function after removing the effects of competing risks. It is inappropriate to use the complement of the Kaplan-Meier (KM) estimator, 1−KM, to summarize the data because this method censors the competing events right after they happen and therefore it assumes "independent censoring" for the competing events. While in most cases, competing risks can not be treated as "independent censoring" since there is usually some dependence structure between the main event and the competing events. We can not obtain the net survival probability for the event of interest given the data since we only observe the failure time of competing events and miss the potential failure time caused by the main event.

According to Peterson (1976), the net survival probability of event 1 should be bounded between $S_T(t) \leq S^1(t) \leq 1 - I^1(t)$ if the risks are dependent. The lower bound is the survival of $min(T_1, T_2)$. In other words, it is the overall survival function, which treats both event 1 and event 2 as events. The upper bound is the complement of the CIF of the event of interest. When the two events (risks) have perfect positive (negative) correlation, the net

survival probability should be exactly equal to the lower (upper) bounds. Researchers (Klein and Moeschberger,1988; Zheng and Klein, 1994) have tightened the bounds and performed inference tests by assuming a dependence structure for the joint distribution of the competing risks. Therefore, the dependence structure of the events is the key point for determining the net probability. However, the dependence structure is "unidentifiable" based on the observed survival time and event.

To estimate the survival function, some early works focused on constructing the censoring models (Williams, Lagakos, 1977; Lagakos, Williams, 1978; Link WA, 1984). Williams and Lagakos's "cone-class" censoring models need the assumption of the distribution of the survival function involved with parameter estimation from the likelihood, while Link's frailty censoring model needs the frailty assumption. Recently, "auxiliary variables" have been introduced to recover missing information due to the dependent censoring. When auxiliary variables are available, Robins and Rotnitzky (1992), Robins (1993), Satten et al (2001), and Robins and Finkelstein (2001) proposed useful techniques to incorporate these variables and recover the information of the dependence structures. By utilizing the auxiliary variables, the survival estimates can be adjusted and imposed efficiently. However, these approaches do not distinguish the dependent and independent censoring and therefore the following estimates are strongly dependent on the sufficiency of the auxiliary variables.

In this paper, we propose new approaches to obtain the net survival probability estimators under various negative/positive correlation structures. First, for data with perfect negative correlation, we proposed an adjusted survival estimator, which is equivalent to 1-CIF. This adjusted survival function was constructed based on the Kaplan-Meier estimator by utilizing the Inverse Probability of Censoring Weight (IPCW) to adjust the potential censoring status for event 2 persons after their event 2 time. Second, after introducing the bounds for negative correlation, the adjusted survival function for negative correlation was proposed by using the adjusted number at death. Specifically, for the event 2 subjects, their censoring status was adjusted by assigning a new weight function which is a linear function of the IPCW and an index of auxiliary variables. Third, by introducing the bounds for positive correlation, the adjusted survival function for positive correlation was constructed using the adjusted number of events. This adjusted number of events takes account of the potential occurrence

of event 1 from event 2 persons. And the auxiliary variables were incorporated to recover this information. The properties of these adjusted survival functions have been derived. Modified log rank tests are proposed to test the equality of two adjusted survival functions.

5.2 METHOD

5.2.1 Notations

To simplify, we call event 1 the event of interest; and call event 2 for the other event(s) due to competing risk(s). Let $(T_i, \delta_i, \epsilon_i)$ denote an independent sample of right-censored survival data; where T_i is the possibly right censored event time; δ_i is the censoring indicator, $\delta_i = 0$ if T_i is censored and $\delta_i = 1$ if T_i corresponds to an event; ϵ_i is the event indicator, $\epsilon_i = 1$ means the ith patient has event 1; $\epsilon_i = 2$ means the ith patient has event 2. Also, suppose in this data that there are D distinct event 1 times $t_1 < t_2 < \cdots < t_D$ and that the total number of patients is Y at the beginning of the study. Let e_j be the number of deaths (observed, from event 1 persons), and e'_{2j} be the number of deaths (unobserved, from event 2 persons) at time t_j; let r_j^* be the number of observations censored (unobserved, from event 2 persons) at time t_j.

5.2.2 Inverse Probability of Censoring Weight (IPCW)

The inverse probability of censoring weight (IPCW) was first presented by Robins and Rotnitzky (1992). Gray and Fine (1999) adapted their idea and created a function of IPCW for the event 2 patients when they constructed the partial likelihood of the sub-distribution (CIF). As we mentioned before, when event 1 and event 2 have perfect negative correlation, which means event 2 subjects have no chance to experience event 1 in the unknown future time, the net survival probability is equal to 1-CIF.

For the events with perfect negative correlation, we can impute the censoring status for the event 2 subjects using the inverse censoring probability. Our basic idea is:

1. Subjects who experience event 2, are treated as if they will experience event 1 at time $t = \infty$. Thus, with respect to event 1, the problem is transformed to a univariate case, where it is assumed that every subject will eventually experience an event 1.
2. For the ith person who experienced event 2 at time $T_i < t$, it is unknown whether this person would be censored at the analysis time t. But we know that this person is still free of event 1 at time T_i. For any $t > T_i$, we assign a weight $0 < w_i < 1$. And the weight is a function of an inverse probability of censoring.

   Our IPCW weight for the event 2 person is given by

$$w_i(t) = \begin{cases} 0 & \text{if } t \leq T_i \\ 1 - \frac{P(C>t)}{P(C>T_i)} & \text{if } t > T_i \end{cases}.$$

This function is increasing as analysis time, t is increasing. Specifically, this weight is a "censoring weight", it is an estimated probability of the potential censoring status for an event 2 subject at each analysis time t. For an event 2 patient, his censoring weight is 0 before he experiences event, so he is in the risk set with a weight of 1; after the occurrence of event 2, he is in the risk set with a weight less than 1 depending on the censoring distribution and his event 2 time. Satten and Datta (2001) showed that the Kaplan-Meier estimator could be used as an inverse probability of censoring weighted average. Thus, to get the weight, we can use the Kaplan-Meier method to obtain the distribution of censoring, then calculate the weight at each time. Specifically, we can treat censoring as an event, while event 1 and 2 as censoring, then we perform the regular Kaplan-Meier analysis using a standard statistical package.

5.2.3 Adjusted Survival Function for Perfect Negative Correlation

We start from the situation with perfect negative correlation. In this case, event 2 persons have no chance of getting event 1 in the unknown future time. Here we propose an adjusted survival estimator which is extended from the Kaplan-Meier framework. In our approach, we focus on interpreting the effect of event 1 after we take account of the event 2 patients. In this adjusted survival function, the numerator is the number of deaths (event 1) at each time t_j, this should have no unobserved event from event 2 persons; the denominator is the

adjusted number at risk, we adjust the weight at risk set for event 2 persons using our IPCW weight.

Our proposed adjusted survival estimator of event 1 is given by

$$\hat{S}'^1(t) = \prod_{j=1}^{D}(1 - \frac{e_j}{Y'_{j-1}}), \tag{5.1}$$

where $Y'_j = Y - \sum_{k=1}^{j}(e_k + c_k) - r'_j$.

Using the index function, we can rewrite the e_k, c_k, r_k as follows

$$\begin{aligned} e_k &= \textstyle\sum_{i:T_i=t_k} \delta_i \cdot I(\epsilon_i = 1) \\ r'_j &= \textstyle\sum_{i:T_i<t_j} w_i(t_j)\delta_i \cdot I(\epsilon_i = 2) \\ c_k &= \textstyle\sum_{i:t_{k-1}\le T_i<t_k} I(\delta_i = 0), \end{aligned}$$

where $w_i(t)$ is the weight function for the ith individual who experienced event "2" and $0 < w_i(t) < 1$. It is a time dependent function of T_i and t_j. So for each event 2 subject, there can be a different w_i at different analysis times.

This adjusted survival function is a net survival estimator for events with perfect negative correlation. For event 2 persons, their censoring starts to happen right after their event 2 time, and the censoring probability follows our IPCW over time. Their weight in the risk set is decreasing as time increases after the event 2 time.

The adjusted survival function we proposed is an estimated "net survival" under the assumption of perfect negative correlated risks. Therefore, it should be equivalent to 1−CIF. Here is the short proof to show this equivalence. The net survival of event 1 can be written as

$$\hat{S}^1(t) = \prod_{j=1}^{D}(1 - \frac{e_j + e'_{2j}}{Y - \sum_{k=1}^{j}(e_k + c_k) - r^*_j}). \tag{5.2}$$

If the risks are perfectly negatively correlated,

$$1 - I^1(t) = S^1(t),$$

and $e'_{2j} = 0$. Thus,

$$\begin{aligned}\hat{S}^1(t) &= \prod_{j=1}^{D}(1 - \frac{e_j + e'_{2j}}{Y - \sum_{k=1}^{j}(e_k + c_k) - r_j^*}) \\ &= \prod_{j=1}^{D}(1 - \frac{e_j}{Y - \sum_{k=1}^{j}(e_k + c_k) - r_j^*}), \end{aligned} \tag{5.3}$$

And this net probability of event 1 has the same formula as our proposed adjusted survival function in (5.1). As far as we can correctly estimate r_j^* for event 2 persons, our adjusted survival function should be equivalent to 1-CIF. As we mentioned previously, the r'_j was constructed via the censoring weight and it should converge to the r_j^* as the sample sizes increase.

5.2.4 Adjusted Survival Function for Various Negative Correlation

The adjusted survival estimator in (5.1) is only valid for the perfect negative correlation. In this section, we try to extend this adjusted survival estimator for events with various negative correlation patterns. As we mentioned before, if the risks are negatively correlated, the bounds can be further written as: $S_{KM}(t) \leq S^1(t) \leq 1 - I^1(t)$. The lower bound is the situation with independent censoring for the event 2 subjects, where the event 2 subjects are censored right at their event 2 time. We can substitute the upper bound with our proposed adjusted survival function. Therefore,

$$S_{KM}(t) \leq S^1(t) \leq S'^1(t).$$

In this case, the upper bound (independent risks) and lower bound (perfect negative correlation) are determined. The underlying true net survival should lie between these two bounds if the risks have unknown negative correlation. No matter what this survival distribution looks like, we always can create an adjusted survival function to approach it. With the sample size large enough, the survival curve should look smooth.

Comparing the upper bound with the lower bound, the only difference comes from the "number at risk". In the lower bound, the risk set is equal to

$$Y_j = Y - \sum_{k=1}^{j}(e_k + c_k) - r_j,$$

where $r_j = \sum_{i:T_i<t_j} \delta_i \cdot I(\epsilon_i = 2)$. While in the upper bound the risk set is equal to

$$Y'_j = Y - \sum_{k=1}^{j}(e_k + c_k) - r'_j,$$

where $r'_j = \sum_{i:T_i<t_j} w_i(t_j)\delta_i \cdot I(\epsilon_i = 2)$.

If the dependence structure exists, the difference must reflect the contribution of event 2 in the risk set. And the "censoring contribution" should between r_j and r'_j. Specifically, for the ith event 2 person, the censoring status at time t ($t > T_i$) should between 1 and $w_i(t)$, and this status is also affected by the dependence structure of the risks. Therefore, for the ith event 2 individual, we construct a specific $w_i^g(t)$

$$w_i^g(t) = p_i * 1 + (1 - p_i) * w_i(t) \qquad \text{if } t > T_i, \tag{5.4}$$

where $p_i = Pr(i\text{th event 2 individual would experience event 1}|\vec{Z}_i)$, $\vec{Z}_i$ is a vector of auxiliary variables. This p_i can be derived from the logistic regression, with $\vec{Z}_i$ as the independent variable. The data for this regression can be obtained from the original dataset excluding the event 2 subjects. We can code event 1 as 1 and censoring as 0. The predicted p_i for the ith event individual can be obtained from the logistic regression equation. It should be noticed that $\vec{Z}_i$ needs to be sufficient to predict the likelihood of event 2 persons experiencing event 1. Specifically, given the auxiliary variables $\vec{Z}_i$, the probability of event 2 subjects experiencing event 1 does not depend on the unobserved times.

To obtain the generalized adjusted survival function for various dependence structures , we replace the $w_i(t)$ with $w_i^g(t)$ in the censoring part of event 2 subjects. Thus,

$$r'_j = \sum_{i:T_i<t_j} w_i^g(t_j)\delta_i \cdot I(\epsilon_i = 2), \tag{5.5}$$

and following equation (5.1), we calculate the survival function.

If for each event 2 individual, $p_i = 1$ and $w_i^g(t) = 1$, our adjusted survival function approaches the lower bound $\hat{S}'^1(t) = S_{KM}$. If for each event 2 individual, $p_i = 0$ and $w_i^g(t) = w_i(t)$, our adjusted survival function converges to the upper bound.

In our generalized adjusted survival function, the "pseudo-like" censoring part from event 2 subjects is not a traditional censoring status for these patients. In traditional Kaplan-Meier survival, the censoring part does not have an event, these persons are lost to follow-up randomly. While in our approach, since we incorporated one "pseudo-like" censoring part to adjust both the potential death or censoring, this part has to include both the potential death and censoring information. Therefore, this penalty part is not the typical part of censoring anymore, it is a combination factor depending on the missing death or censoring. In our approach, we actually constructed hypothetical survival curve to approach the true underlying survival. The form of $w_i^g(t)$ is a linear combination of p_i and $w_i(t)$. The impact of a different choice of this form should be further investigated via simulation studies.

5.2.5 Adjusted Survival Function for Various Positive Correlation

When the risks are positively correlated, the upper and lower bound of the net survival of event 1 is:

$$S_T(t) \leq S^1(t) \leq S_{KM}(t).$$

Therefore, for any unknown positive correlation, the net survival should lie between the upper and lower bound. When the risks have perfect positive correlation, the net survival equals the lower bound, which is the overall survival and treats both event 1 and event 2 as an event. When the risks are independent, the net survival reaches the upper bound, the KM survival, which treats event 2 as censoring right at their event 2 time.

Comparing the upper and lower bound, we find that: the only difference of these bounds comes from the difference in the numerator. Here e_{2j} is the number of deaths due to event 2 during time interval t_{j-1} to t_j. The lower bound can be written as

$$\hat{S}_T(t) = \prod_{j=1}^{D}(1 - \frac{e_j + e_{2j}}{Y_{j-1}}), \tag{5.6}$$

while the upper bound is

$$\hat{S}_{KM}(t) = \prod_{j=1}^{D}(1 - \frac{e_j}{Y_{j-1}}), \tag{5.7}$$

where $Y_j = Y - \sum_{k=1}^{j}(e_k + c_k) - r_j$.

Using the index function, we can rewrite the e_k, c_k, and r_k as follows

$$\begin{aligned} e_k &= \textstyle\sum_{i:T_i=t_k} \delta_i \cdot I(\epsilon_i = 1) \\ r_j &= \textstyle\sum_{i:T_i<t_j} \delta_i \cdot I(\epsilon_i = 2) \\ c_k &= \textstyle\sum_{i:t_{k-1}\le T_i<t_k} I(\delta_i = 0) \\ e_{2j} &= \textstyle\sum_{i:t_{j-1}<T_i\le t_j} \delta_i \cdot I(\epsilon_i = 2). \end{aligned}$$

Obviously, the difference between the upper and the lower bound is the number of deaths due to event 2 subjects. With the perfect positive correlation, the event 1 time for the event 2 subjects is equal to their event 2 time. If we want to recover the net survival for data with unknown positive correlation, we can define an adjusted survival function which is similar in form to the upper and lower bound, the key point is to find the adjusted number of deaths due to event 2 subjects (unobserved) at each time interval from t_{j-1} to t_j. And this quantity is unidentifiable given the observed survival time and event. Similarly as in the previous part, we need to borrow information from the auxiliary variables. If, given the auxiliary variables, the hazard of event 1 does not depend on the unobserved censoring or event, then these auxiliary variables are sufficient to recover the missing event 1 information due to competing risks.

Suppose the ith event 2 person experienced event 2 during time t_{j-1} to t_j. From the upper and lower bound, we can see that this person is out of the risk set, but whether he experiences event 1 or censoring is not certain. Therefore, we can use the auxiliary variable to predict the potential probability of experiencing event 1 and the adjusted number of event 1 due to event 2 persons has the form

$$e'_{2j} = \sum_{i:t_{j-1}<T_i\le t_j} p_i\delta_i \cdot I(\epsilon_i = 2) \tag{5.8}$$

where $p_i = Pr(\text{ith event 2 individual would experience event 1}|\vec{Z}_i)$ and $\vec{Z}_i$ is a vector of auxiliary variables. This p_i has the same definition as the one we used in the negative correlation

case and can be obtained via the approach we described previously. The adjusted survival function for positive correlation can be obtained from

$$\hat{S}'^1(t) = \prod_{j=1}^{D}(1 - \frac{e_j + e'_{2j}}{Y_{j-1}}). \tag{5.9}$$

If $p_i = 1$ for each event 2 individual, our adjusted survival function approaches the lower bound $\hat{S}'^1(t) = S_T(t)$. If $p_i = 0$ for each event 2 individual, our adjusted survival function converges to the upper bound $\hat{S}'^1(t) = S_{KM}(t)$.

5.3 PROPERTIES

5.3.1 Mean and Variance

In this section, we start to derive the properties for the adjusted survival function with negative correlations.

We define $s'_j = S'^1(t_j)/S'^1(t_{j-1})$, and its adjusted estimator is $\widehat{s}'_j = 1 - \frac{e_j}{Y'_{j-1}}$ and E_j is a conditional expectation given information up to time t_j. First, we show that $\widehat{s}'_j$ is unbiased, and then show that $\widehat{S}'^1(t_j)$ is unbiased by means of successive conditional expectation, that is,

$$\begin{aligned} E_j\{1 - \widehat{s}'_j\} &= E_j(\frac{e_j}{Y'_{j-1}}|\text{information up to time } t_j) \\ &= \frac{(1 - s'_j) \cdot Y'_{j-1}}{Y'_{j-1}} \\ &= 1 - s'_j. \end{aligned}$$

Here the expected number of deaths at time t_j is equal to the probability of death multiplied by the number at risk and we plug in the new adjusted number at risk instead of the usual

one. Following the above notation, by means of successive expectations, we obtain

$$\begin{aligned} E\{\widehat{S}'^1(t_j)\} &= E\{\widehat{s}'_1 \cdots \widehat{s}'_2 \cdot \widehat{s}'_{j-1} \cdot E(\widehat{s}'_j)\} \\ &= E\{\widehat{s}'_1 \cdot \widehat{s}'_2 \cdots \widehat{s}'_{j-1}\} \cdot s'_j \\ &= E\{\widehat{s}'_1 \cdot \widehat{s}'_2 \cdots \widehat{s}'_{j-2}\} \cdot E(\widehat{s}'_{j-1}) \cdot s'_j \\ &= s'_1 \cdot s'_2 \cdots s'_j \\ &= S'^1(t_j). \end{aligned}$$

From the above proof, we can show that the adjusted survival function can be unbiased if the weight is proper. From the successive expectations, the estimator is appropriate whenever t_j is less than the maximum event 1 time.

Now we can derive the variance formula of this adjusted survival function

$$\begin{aligned} E\{\hat{S}'^1(t)\}^2 &= \prod_{j:T_j \le t} E_j(\hat{S}'^1_j)^2 \\ &= \prod_{j:T_j \le t} \{S'^1_j + Var_j(\hat{S}'^1_j)\} \\ &= \prod_{j:T_j \le t} \{(S'^1_j)^2 + \frac{S'^1_j(1 - S'^1_j)}{Y'_j}\} \\ &= (S'^1_1)^2 ... (S'^1_k)^2 \prod_{j:T_j \le t} \{1 + \frac{1 - S'^1_j}{Y'_j S'^1_j}\} \\ &= \{S'^1(t)\}^2 \prod_{j:T_j \le t} (1 + \frac{1 - S'^1_j}{Y'_j S'^1_j}). \end{aligned}$$

By the definition of the variance, we can get

$$\begin{aligned} Var\{\hat{S}'^1(t)\} &= E\{\hat{S}'^1(t)\}^2 - [E\{\hat{S}'^1(t)\}]^2 \\ &= S'^1(t)^2 \{ \prod_{j:T_j \le t} (1 + \frac{1 - s'_j}{Y'_j s'_j}) - 1\}. \end{aligned} \tag{5.10}$$

Following the same approach, we also can show that our adjusted survival for positive correlation is unbiased. And the variance formula can also be derived in the same way except Y' is replaced by Y in equation (5.10).

5.3.2 Confidence Interval

The $(1-\alpha)\times 100\%$ pointwise confidence interval for an adjusted survival function is $\hat{S}'^{r}(t) \pm Z_{1-\alpha/2} \cdot \widehat{Var}[\hat{S}'^{r}(t)]^{1/2}$.

5.4 MLE OF PSEUDO-LIKELIHOOD

For right-censored data in the absence of competing risks, we can write the likelihood with respect to the survival function. For patients who experience the event, we use the density function; for the patients who are censored, we use the survival function. With right-censored data in the presence of competing risks, we can write a "pseudo-likelihood" with respect to the adjusted survival function for event 1. In this case, we still use the density function for subjects who experience event 1, and the survival function for "true censoring". But for the subjects who experience event 2, their censoring times are unknown, and the event 1 times are unobserved. For data with negative correlation, we can treat event 2 subjects as being partially censored at each time point (event 1 time) with weight $w_i^g(t)$. Similarly, for data with positive correlation, we can treat event 2 subjects as a partial event with weight p_i and as partially censored with weight $1-p_i$.

5.4.1 MLE of Pseudo-likelihood for Negative Correlation

Let $t_1 < t_2 < \cdots < t_k$, denote the distinct event 1 times; λ_j denote the number of true censored times in $[t_j, t_{j+1})$; r_j denote the number of event 2 subjects during $[t_j, t_{j+1})$; w_i^g denote vector of the weights for the ith person. For the ith person, $w_i^g(t_j)$ is a different weight at each time point t_j. A pseudo-likelihood with respect to $S(t)$ (here $S(t)$ refers to the adjusted survival $S'(t)$) is then defined as:

$$\prod_{i=1}^{n} [S_i(T_i - 0) - S_i(T_i)]^{\delta_i I(\epsilon_i = 1)} [S_i(T_i)]^{(1-\delta_i)} [S_i(T_i)]^{\delta_i I(\epsilon_i = 2) w_i^g}. \tag{5.11}$$

We can rewrite the pseudo-likelihood in terms of Kaplan-Meier as

$$\begin{aligned} L\{S(t)\} &= \prod_{i=1}^{\lambda_0}\{S(L_i^{(0)})\} \times \prod_{i=1}^{r_0}\{S(L_i^{(0)})\}^{w_i^g(t_1)\delta_i I(\epsilon_i=2)} \times \{S(t_1-0)-S(t_1)\}^{\sum_{i:T_i=t_1}\delta_i I(\epsilon_i=1)} \\ &\times \prod_{i=1}^{\lambda_1}\{S(L_i^{(1)})\} \times \prod_{i=1}^{r_0+r_1} \times\{S(L_i^{(1)})\}^{w_i^g(t_2)\delta_i I(\epsilon_i=2)} \times \{S(t_2-0)-S(t_2)\}^{\sum_{i:T_i=t_2}\delta_i I(\epsilon_i=1)} \\ &\cdots \times\{S(t_k-0)-S(t_k)\}^{\sum_{i:T_i=t_k}\delta_i I(\epsilon_i=1)} \times \prod_{i=1}^{\lambda_k}\{S(L_i^{(k)})\} \times \prod_{i=1}^{r_0+r_1+\cdots+r_k}\{S(L_i^{(k)})\}^{w_i^g(t_{k+1})\delta_i I(\epsilon_i=2)}. \end{aligned}$$

For right-censored data, we know that

$$S(L_i^{(0)}) = S(t_1-0) = 1 \text{ and } S(t_j) = S(L_i^{(j)}) = S(t_{j+1}-0).$$

So (5.11) can be rewritten as

$$\prod_{j=1}^{k}(S_{j-1}-S_j)^{\sum_{i:T_i=t_j}\delta_i I(\epsilon_i=1)} \times (S_j)^{\sum_{i=1}^{\lambda_j}} \times (S_j)^{\sum_{i:T_i<t_j} w_i^g(t_j)\delta_i I(\epsilon_i=2)}. \tag{5.12}$$

From the property of the Kaplan-Meier estimator, let $s_j = S_j/S_{j-1}, S_j = s_1.s_2...s_j$, so $S_{j-1}-S_j = s_1...s_{j-1}(1-s_j)$ and (5.12) can be written as

$$\begin{aligned} Ł(S_j) &= \prod_{j=1}^{k}(1-s_j)^{\sum_{i:T_i=t_j}\delta_i I(\epsilon_i=1)} \cdot (s_j)^{\sum_{l=j+1}^{k}\{\sum_{i=1}^{\lambda_l}+\sum_{i:T_i=t_l}\delta_i I(\epsilon_i=1)+\sum_{i:T_i<t_l} w_i^g(t_l)\delta_i I(\epsilon_i=2)\}} \\ &\cdot (s_j)^{\sum_{i=1}^{\lambda_j}+\sum_{i:T_i<t_j} w_i^g(t_j)\delta_i I(\epsilon_i=2)}. \end{aligned} \tag{5.13}$$

The above equation can be further simplified as

$$\prod_{j=1}^{k}(1-s_j)^{\sum_{i:T_i=t_j}\delta_i I(\epsilon_i=1)} \cdot (s_j)^{\sum_{i:T_i\geq t_j}-\sum_{i:T_i=t_j}\delta_i I(\epsilon_i=1)}, \tag{5.14}$$

where we utilize the fact that

$$\textstyle\sum_{i:T_i\geq t_j} = \sum_{l=j}^{k}\{\sum_{i=1}^{\lambda_l}+\sum_{i:T_i=t_l}\delta_i I(\epsilon_i=1)+\sum_{i:T_i<t_l} w_i^g(t_l)\delta_i I(\epsilon_i=2)\}.$$

In (5.14), we take the log likelihood first, and then compute the first derivative with respect to s_j . So each s_j is maximized individually by the binomial estimate

$$\hat{s}_j = \frac{\sum_{i:T_i\geq t_j}-\sum_{i:T_i=t_j}\delta_i I(\epsilon_i=1)}{\sum_{i:T_i\geq t_j}} = 1-\frac{e_j}{Y'_{j-1}}.$$

Therefore, the survival function is estimated by

$$\widehat{S}_j'^1(t) = \begin{cases} 1 & \text{if } t < t_1 \\ \prod_{t_j \le t}[1 - \frac{e_j}{Y'_{j-1}}] & \text{if } t_1 \le t \end{cases}. \tag{5.15}$$

Thus, the adjusted survival function is the MLE of the pseudo-likelihood.

5.4.2 MLE of Pseudo-likelihood for Positive Correlation

For data with positive correlation, let p_i denote the weight for the ith person. A pseudo-likelihood with respect to $S(t)$ (here $S(t)$ refers to the adjusted survival $S'(t)$) is defined as:

$$\prod_{i=1}^{n}[S_i(T_i-0)-S_i(T_i)]^{\delta_i I(\epsilon_i=1)}[S_i(T_i)]^{(1-\delta_i)}[S_i(T_i-0)-S_i(T_i)]^{\delta_i I(\epsilon_i=2)p_i}[S_i(T_i)]^{\delta_i I(\epsilon_i=2)(1-p_i)}, \tag{5.16}$$

Following the approach of the previous proof, we can rewrite this pseudo-likelihood into the Kaplan-Meier setting, and then further simplify it as

$$\prod_{j=1}^{k}(S_{j-1}-S_j)^{\sum_{i:T_i=t_j}\delta_i I(\epsilon_i=1)+\sum_{i:t_{j-1}<T_i\le t_j}\delta_i I(\epsilon_i=2)p_i} \times (S_j)^{\sum_{i=1}^{\lambda_j}} \times (S_j)^{\sum_{i:t_{j-1}<T_i\le t_j}\delta_i I(\epsilon_i=2)(1-p_i)}.$$

Utilizing the properties of the Kaplan-Meier estimator, the above equation is equal to

$$\prod_{j=1}^{k}(1-s_j)^{\sum_{i:T_i=t_j}\delta_i I(\epsilon_i=1)+\sum_{i:t_{j-1}<T_i\le t_j}\delta_i I(\epsilon_i=2)p_i}(s_j)^{\sum_{i:T_i\ge t_j}-\sum_{i:T_i=t_j}\delta_i I(\epsilon_i=1)-\sum_{i:t_{j-1}<T_i\le t_j}\delta_i I(\epsilon_i=2)p_i}, \tag{5.17}$$

where we utilize the fact that

$$\textstyle\sum_{i:T_i\ge t_j} = \sum_{l=j}^{k}\{\sum_{i=1}^{\lambda_l} + \sum_{i:T_i=t_l}\delta_i I(\epsilon_i=1) + \sum_{i:t_{l-1}<T_i\le t_l}\delta_i I(\epsilon_i=2)\}.$$

In the equation (5.17), we compute the log likelihood first and then obtain the first derivative with respect to s_j. The MLE of s_j then has the form

$$\hat{s}_j = \frac{\sum_{i:T_i\ge t_j} - \sum_{i:T_i=t_j}\delta_i I(\epsilon_i=1) - \sum_{i:t_{j-1}<T_i\le t_j}\delta_i I(\epsilon_i=2)p_i}{\sum_{i:T_i\ge t_j}} = 1 - \frac{e_j + e'_{2j}}{Y_{j-1}}. \tag{5.18}$$

Therefore, the adjusted survival function for the positive correlation $\hat{S}_j'^1(t)$ is the MLE of the pseudo-likelihood.

5.5 INFERENCE TEST

When we try to compare the summary function of the cause of interest between two groups (for example, the treatment group vs. the control group in a cancer study), our primary goal is to compare the treatment effect on cancer survival and we do not care about competing risks, such as death due to old age and other causes. Thus, we can use our adjusted survival function for testing the differences between the two groups. We know that the log rank test is the most powerful test for survival data with proportional hazards. In our case, we can adapt the idea of the log rank test since the adjusted survival function ($S'^r(t)$) can be treated as the traditional survival function. Therefore, in our modified log rank test, we can test the equality of two $S'^r(t)$s.

Consider the hypothesis

$$H_0 : S_1'^r(t) = S_2'^r(t) \qquad \text{for all } t \leq \tau$$
$$H_a : S_1'^r(t) \neq S_2'^r(t) \qquad \text{for some } t \leq \tau,$$

where τ is the largest time at which both groups have at least one subject at risk.

Here we start with the inference test for negative correlation. Let $d_j = d_{j1} + d_{j2}$ and $Y'_j = Y'_{j1} + Y'_{j2}$ be the number of deaths and the adjusted number at risk at time t_j, $j = 1, 2 \cdots, D$. Based on Klein and Moeschberger (2003), our modified log-rank test is

$$G = \sum_{j=1}^{D} \{d_{j1} - Y'_{j1}(\frac{d_j}{Y'_j})\}. \tag{5.19}$$

Note that at each time, t_j, the test statistic is a multinomial random variable with parameters d_j and $p_j = \frac{Y'_{j1}}{Y'_j}$. In other words, conditional on d_j, d_{j1} has a hypergeometric distribution. So the variance of G is

$$Var(G) = \sum_{j=1}^{D} \frac{Y'_{j1}}{Y'_j}(1 - \frac{Y'_{j1}}{Y'_j})(\frac{Y'_j - d_j}{Y'_j - 1})d_j. \tag{5.20}$$

The adjusted log-rank test is proposed as $Z = \frac{G}{\sqrt{Var(G)}}$, which has a standard normal distribution for large samples when H_0 is true. Thus, H_0 is rejected if $|Z| \geq Z_{\alpha/2}$ at an α level, where $Z_{\alpha/2}$ is the critical value in the normal distribution.

For data with positive correlation, our modified log-rank test has this form

$$G = \sum_{j=1}^{D} \{d'_{j1} - Y_{j1}(\frac{d'_j}{Y_j})\}, \tag{5.21}$$

and the variance of G is

$$Var(G) = \sum_{j=1}^{D} \frac{Y_{j1}}{Y_j}(1 - \frac{Y_{j1}}{Y_j})(\frac{Y_j - d'_j}{Y_j - 1})d'_j. \tag{5.22}$$

It should be noted that the variance formula we give is problematic due to the fact of Y'_j, Y'_{j1}, d'_j, and d'_{j1} may not be integers. In the usual log rank test, both the number at risk and the number of deaths are integers. In our modified log rank test, we use the "adjusted" number at risk or "adjusted" number of deaths, and these quantities are fractional numbers. The impact of these non-integer items on the variance formula needs to be further studied.

5.6 SIMULATION STUDIES

Because of the unknown features of the auxiliary variables, it is difficult to perform simulation studies for the adjusted survival function with various correlations. Here we only evaluate the performance of the adjusted survival function under the case of perfect negative correlation.

5.6.1 Performance of the Adjusted Survival Estimators

A simulation study was carried out to examine the performance of the estimators of the mean, variance and associated confidence intervals in simulated samples with limited size. The dependence structure between event 1 and event 2 was modeled as: (1) Let T_1 and T_2 denote two lognormally distributed random variables, where $T_j = e^{Y_j}$ and (Y_1, Y_2) follows a bivariate normal distribution. Let u_j and σ_j denote the mean and standard deviation of Y_j; and let $u_1 = u_2 = 0$, $\sigma_1 = \sigma_2 = 1$. (2) We transformed the correlation ρ between Y_1 and Y_2 into the Spearman correlation $r = (6/\pi)arcsin(\rho/2)$, such that this number was both the Spearman correlation between Y_1 and Y_2 as well as between T_1 and T_2; (3) Three different correlations were studied, $\rho = 0, 0.5, -0.5$. Independent censoring was generated from a

uniform distribution, and two censoring rates were used, 0.15, 0.30. The sample size was set at 100 and 500 samples were generated. Three time points were used, $t_0 = 0.5, 1.0, 1.5$.

We calculated the mean, variance, and 95% confidence interval from the sample, and compared them with the values we obtained from our formula. The results are displayed in Table 5. Under different correlation structures, the estimator of adjusted survival seems unbiased at each time point for both censoring rates. Although the variance estimator appears to be slightly biased for some of the cases, the 95% confidence interval calculated from our estimator had good coverage for practical purposes.

Table 5. Performance of the Adjusted Survival Estimators

Correlation ρ	Censoring rate	Time point t_0	$S'^1(t_0)$	Mean $(\hat{S}'^1(t_0))$	var $(\hat{S}'^1(t_0))$	Mean $(\widehat{var}(\hat{S}'^1(t_0))$	Coverage of 95% C.I.
0	0.15	0.5	0.797	0.800	0.00154	0.00174	95%
		1.0	0.630	0.633	0.00266	0.00261	95%
		1.5	0.565	0.567	0.00314	0.00297	93%
	0.30	0.5	0.822	0.820	0.00159	0.00178	95%
		1.0	0.652	0.654	0.00279	0.00302	95%
		1.5	0.565	0.577	0.00293	0.00345	94%
0.5	0.15	0.5	0.831	0.830	0.00143	0.00152	95%
		1.0	0.679	0.680	0.00234	0.00242	95%
		1.5	0.605	0.604	0.00270	0.00272	95%
	0.30	0.5	0.845	0.847	0.00143	0.00151	95%
		1.0	0.695	0.701	0.00280	0.00269	91%
		1.5	0.620	0.622	0.00333	0.00320	93%
-0.5	0.15	0.5	0.771	0.779	0.00184	0.00188	91%
		1.0	0.591	0.593	0.00287	0.00277	93%
		1.5	0.526	0.528	0.00303	0.00294	93%
	0.30	0.5	0.786	0.795	0.00220	0.00200	88%
		1.0	0.594	0.605	0.00347	0.00330	90%
		1.5	0.522	0.536	0.00400	0.00365	87%

5.6.2 Equivalence of Adjusted Survival Function and 1-CIF

In this section, we performed simulation studies to compare our adjusted survival function to 1-CIF. Samples are generated as the previous simulation study and a sample size of 1000 was used. Three different correlations were studied, $\rho = 0, 0.5, -0.5$, and the censoring rate was 0.15, 0.30.

The simulation results are shown in Fig 8-10. The adjusted survival functions are close to 1-CIF for different correlations and censoring rates, indicating that the two functions are similar under these scenarios.

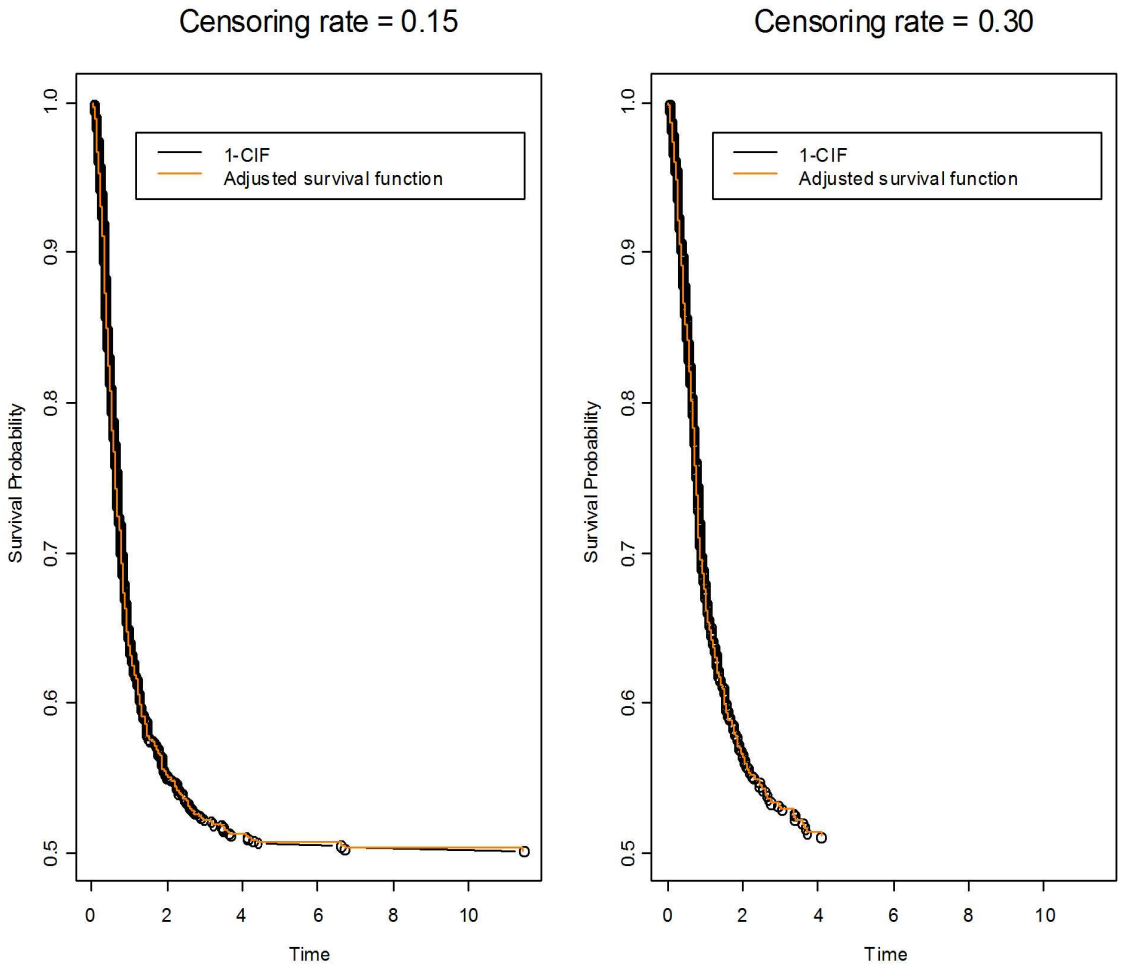

Figure 8. Comparison of the estimated adjusted survival function and the 1-CIF with correlation rate $\rho = 0$.

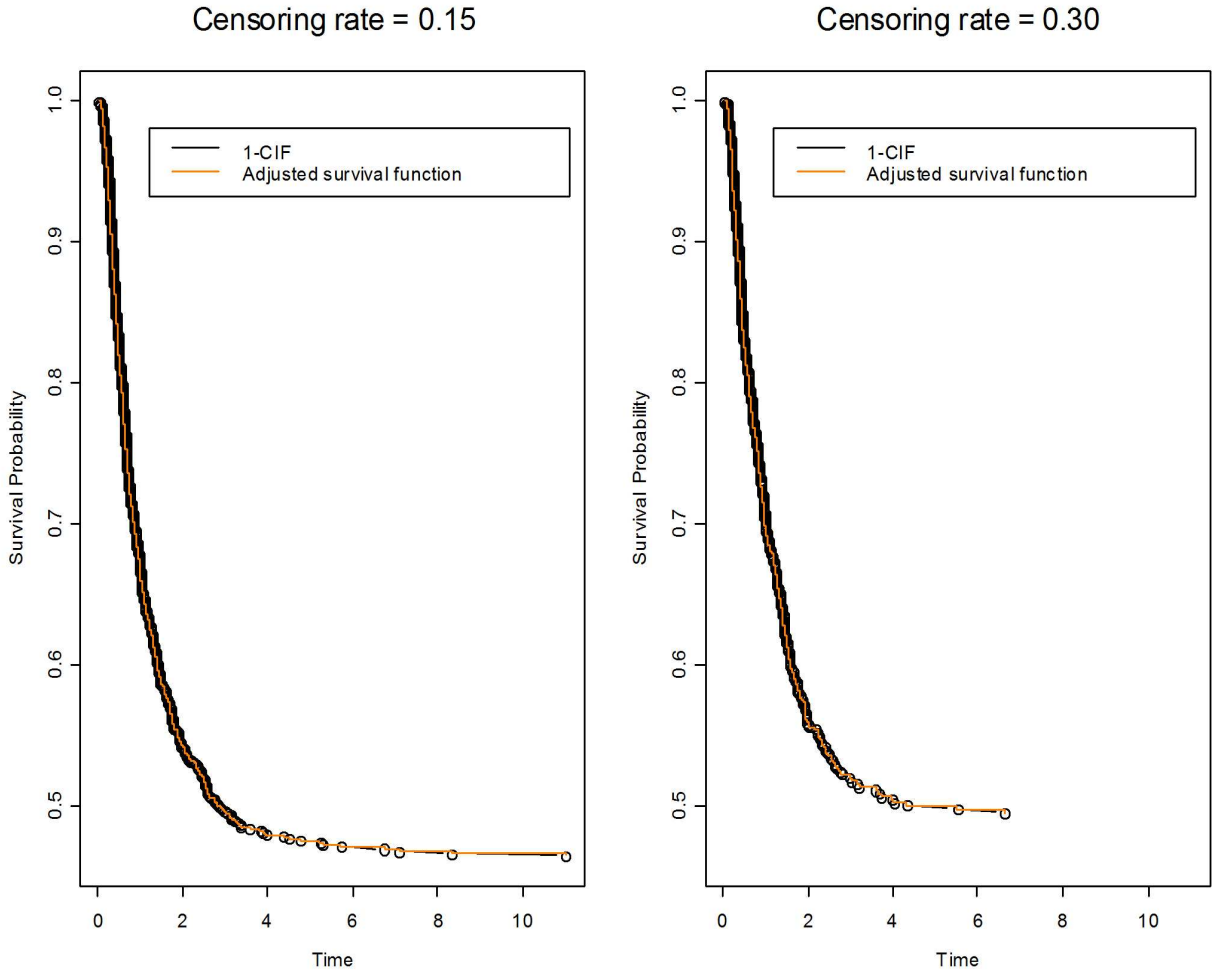

Figure 9. Comparison of the estimated adjusted survival function and the 1-CIF with correlation rate $\rho = 0.5$.

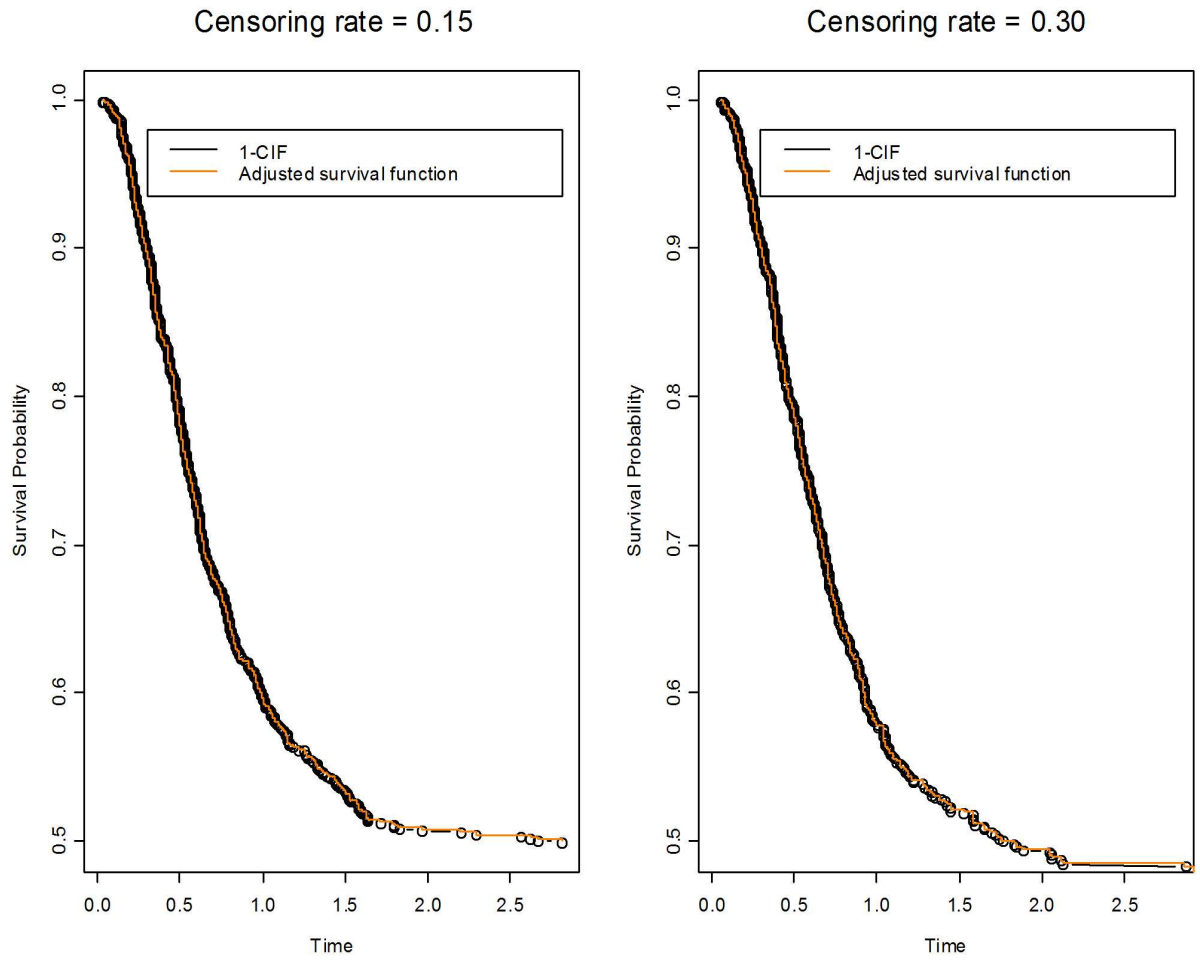

Figure 10. Comparison of the estimated adjusted survival function and the 1-CIF with correlation rate $\rho = -0.5$.

5.7 ANALYSIS OF EMPHYSEMA DATA

The primary goal of the NETT study is to determine the role, safety, and effectiveness of bilateral lung volume reduction surgery (LVRS) for the treatment of emphysema. In the NETT trial, eligible patients are randomized into two groups: a group using medical treatment only (N=610) and a group undergoing surgery along with medical treatment (N=608). Comparing the two groups, no significant survival benefit was shown in the LVRS with medical treatment group. However, the health-related quality of life (HRQL) tends to be improved for the patients in the surgery group, and more patients experienced HRQL improvement in the LVRS group than in the medical treatment only group (Figure 11). As we mentioned earlier, the main event of interest is improvement in HRQL, while death before reaching HRQL is a competing risk. The goal is to estimate the probability of HRQL improvement for each of the two groups and to compare the probability difference between the two groups.

Figure 12 shows that the bounds for the probability of HRQL improvement for each of the two groups, and 1- adjusted survival function (perfectly negative) were almost coincident with the CIF curves (the lower bound). If the HRQL improvement and death before reaching HRQL are perfectly negatively related, then the probability of improvement should be the lower bound; if the risks are independent, the probability of improvement is the upper bound (S_{KM}). Assuming the unknown negative dependence between two events because those who died had worse HRQL compared to those who lived, our proposed adjusted survival function was applied to obtain an estimated probability of HRQL improvement after removing the effect of the competing event. The 1- adjusted survival function fell between the upper and lower bound, and our estimated probability of HRQL improvement is significantly different between the two groups (Z=15.2, $P < 0.0001$). Therefore, the LVRS procedure has shown significant benefit to HRQL improvement based on our proposed approach.

To construct the generalized weight for the event 2 subjects ($w^g(t)$), the estimated probability of experiencing event 1 for each event 2 person was obtained from the logistic regression with a dataset excluding the event 2 persons, with three auxiliary variables were included in the regression model: upper lobe (ul), maximum exercise capacity (maxcat), and treatment

group. In the LVRS group, there are 151 event 2 subjects, among them the 39 persons have missing auxiliary variables (ul and maxcat). In the medical treatment only group, 44 out of the 215 event 2 subjects have missing auxiliary variables. For event 2 subjects with missing auxiliary variables, we imputed their probability with the mean predicted probability of the group (0.76 for the LVRS group, 0.34 for the medical treatment only group). The Kaplan-Meier estimator was used as an inverse probability of censoring weighted average.

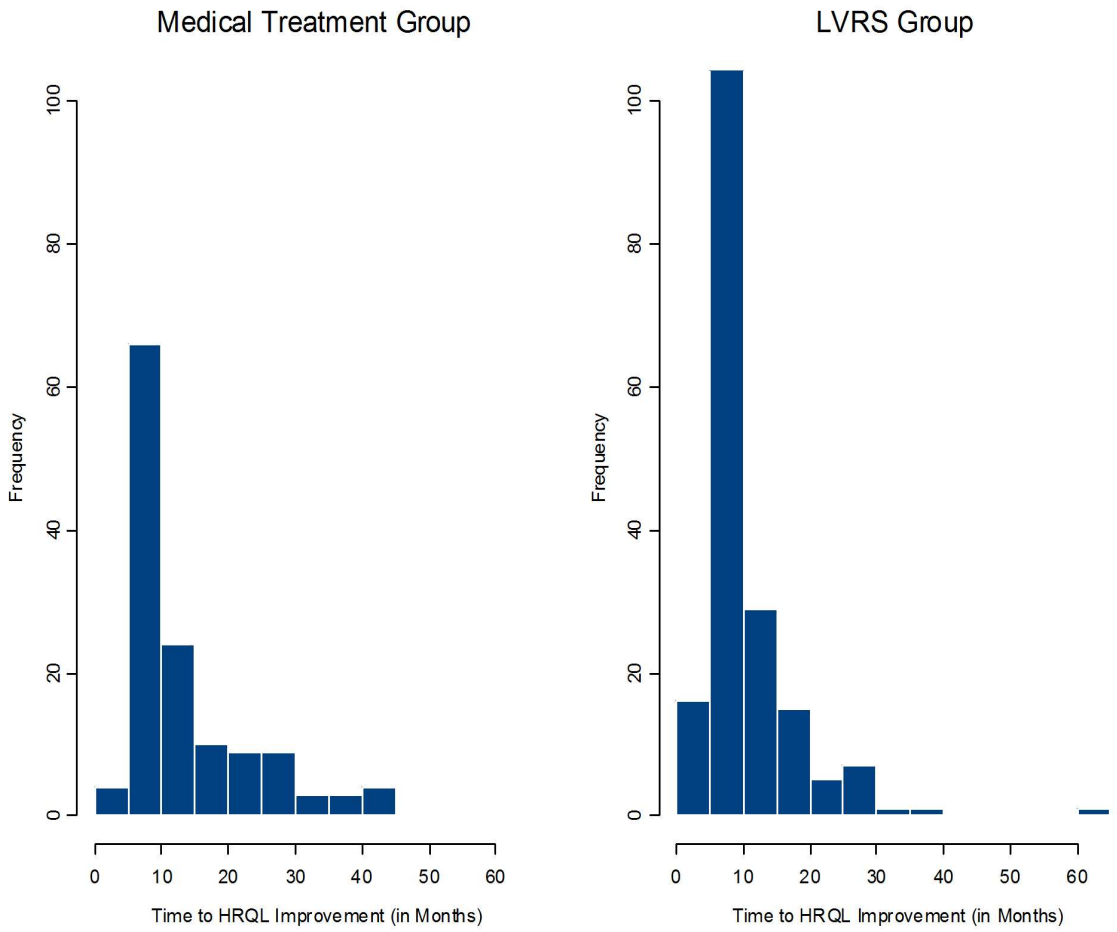

Figure 11. Histogram of time to improvement for each of the two groups.

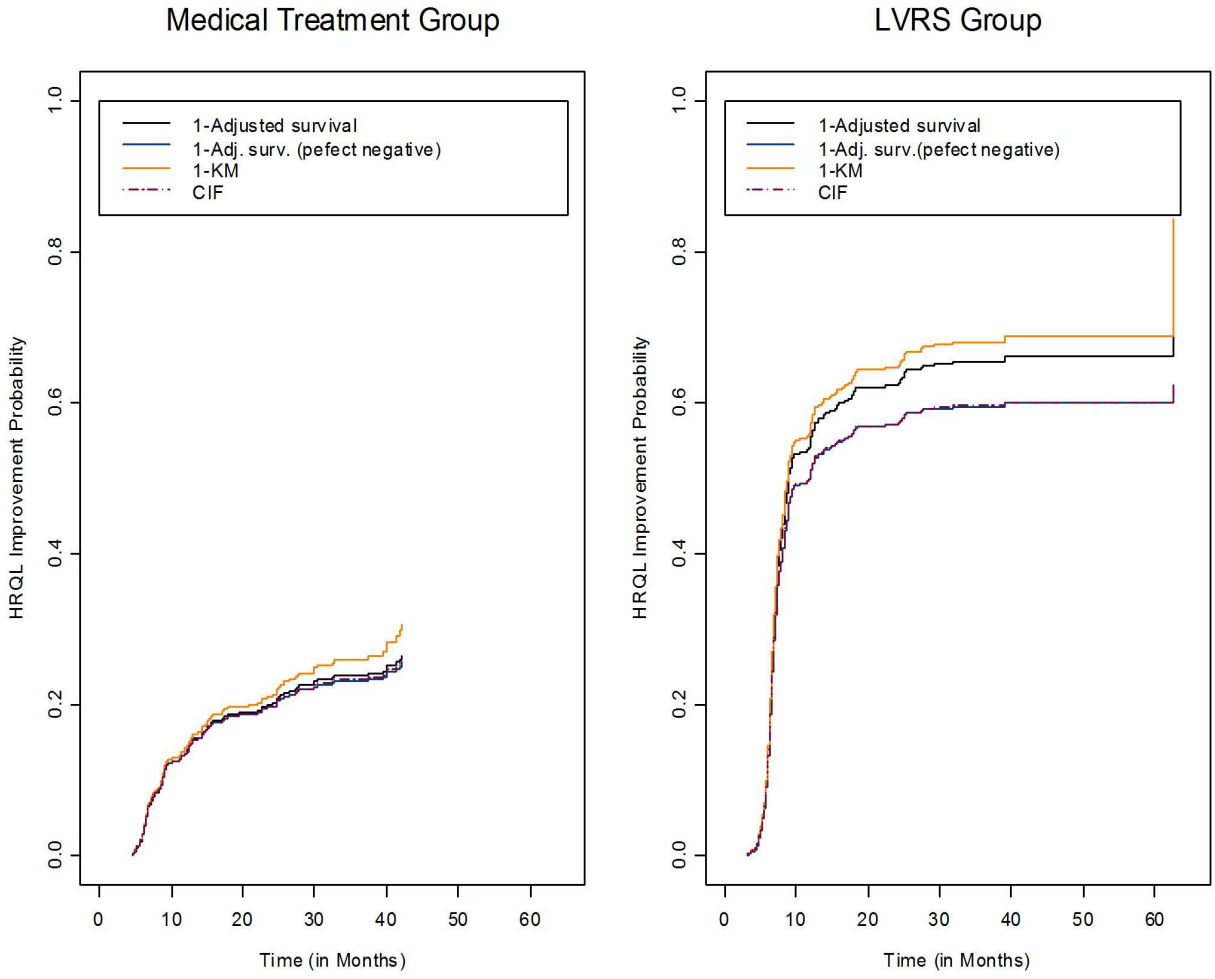

Figure 12. Comparison of the estimated 1-KM, 1-adjusted survival function and the CIF for each of the two groups.

5.8 DISCUSSION

In survival analysis with competing risks, two types of questions are usually of interest: the time to event for each cause or the time to event focused on one primary cause. Currently, the cumulative incidence function (CIF) is the most widely used summary function to address these two types of questions. However, the CIF addresses the marginal failure probability of interest and therefore the resulting function only presents partial information if the failure probabilities of the competing events do not present simultaneously. In many cases, our main interest is focused on one primary event, and the presence of competing risks only make the situation complicated. A single summary function of the "net probability" is more interpretable and appropriate. It is challenging for estimating the survival probability of one event when we lose the information of the potential failure time of the persons experiencing competing events. Without "auxiliary variables", the missing failure time due to the competing events remains unidentifiable given the survival time and event. In order to recover the missing survival information for persons experiencing competing events, we assume that the auxiliary variables are sufficient, which means that given the auxiliary variables, the ith event 2 person's probability of experiencing event 1 at time t does not depend on the potential missing failure time, t_i^*. The discussion of sufficiency of the auxiliary variables can also be found in Satten et al's paper (2001). However, this assumption is untestable and questionable. The violation of this assumption can lead to invalid or biased estimates of the survival probability.

Unlike other net survival probability estimators, our proposed adjusted survival function only targets the event 2 subjects instead of adding weight to all censoring subjects (Robins and Finkelstein, 2000; Grunkemeier et al, 2007). The advantage of our approach is: we treat the true censoring and dependent censoring differently, therefore, only the dependent censoring due to the competing risks needs to be taken care of. Moreover, our adjusted survival functions have properties and inference test similar to the traditional survival functions, leading to easy interpretation and clear understanding. However, for our adjusted survival function, we derived a different formula for negative correlation and positive correlation. Therefore, we have to assume the dependece structures between events before the analysis.

This assumption should be based on the prior knowledge of the events and professional opinions. Mis-specifying the correlation can lead to severe bias in the estimates. The practical procedure for determining the correlation (negative or positive) needs to be studied in the future.

6.0 SUMMARY AND FUTURE WORK

There are two parts for this dissertation: one is to construct an adjusted cumulative incidence function (ACIF) if the data involve unbalanced confounding variables and another is to construct an adjusted survival function if the data involve administrative censoring (independent censoring) and censoring due to competing risks (dependent censoring).

For the first part, the ACIF has been developed and its asymptotic properties have been derived. The ACIF applies the inverse probability weighting technique to the overall survival and cause-specific hazard functions in the naive CIF estimate. The ACIF estimate is proving to be unbiased and the variance formula is derived. An inference procedure to test the equality of two ACIFs is given. The test statistic is based on stochastic ordering of the functions. The performance of the ACIF estimators has been evaluated via simulation studies. In future studies, the performance of the test will be assessed via simulation studies and the test will be generalized from comparing 2 groups to k groups.

For the second part, a series of adjusted survival functions has been developed for estimating the "net" survival probability of interest event. First, for risks with perfect negative correlation, we proposed an adjusted survival function using inverse probability of censoring weight at each time point to remove the effect of dependent censoring due to competing risks. We showed that this adjusted survival function is equivalent to 1-CIF. Second, using Peterson's upper bound and the KM survival estimate, we constructed the bounds for negatively correlated risks. The adjusted survival function was created by using the adjusted number at risk, which incorporated the information recovered from the auxiliary variables for the event 2 subjects and the inverse censoring weight. Third, using Peterson's lower bound and the KM survival estimate, we introduced the bounds for risks with positive correlation. The generalized adjusted survival function was constructed by using the adjusted number

of deaths, which also incorporated the recovered information of the event 2 subjects from auxiliary variables. The properties of these adjusted survival functions have been derived. A generalized log rank test is proposed to test the equality of two adjusted survival functions. The performance of these generalized log rank tests will be studied carefully in the future. Also, the performance of our approach will be assessed via simulation studies and by comparing it with other current methods.

APPENDIX

PROOFS FOR THE ACIF

A.1 MEAN OF THE ACIF

$$
\begin{aligned}
E_j[\hat{I}_k^{rw}(t_j)] &= E_j\left\{\sum_{j'=1}^{j}\hat{\lambda}_k^{rw}(t_{j'})\cdot\hat{S}_k^{w}(t_{j'-1})\right\}\\
&= E_j\left[\sum_{j'=1}^{j}E_{j'-1}\left\{\hat{\lambda}_k^{rw}(t_{j'})\cdot\hat{S}_k^{w}(t_{j'-1})|\text{ information up to time } t_{j'-1}\right\}\right]\\
&= E_j\left[\sum_{j'=1}^{j}\hat{\lambda}_k^{rw}(t_{j'})\cdot E_{j'-1}\left\{\hat{S}_k^{w}(t_{j'-1})|\text{ information up to time } t_{j'-1}\right\}\right]\\
&= E_j\left[\sum_{j'=1}^{j}\hat{\lambda}_k^{rw}(t_{j'})\cdot S_k^{w}(t_{j'-1})\right]\\
&= \sum_{j'=1}^{j}\left[S_k^{w}(t_{j'-1})\cdot E_j\left\{\hat{\lambda}_k^{rw}(t_{j'})\right\}\right]\\
&= \sum_{j'=1}^{j}\left[S_k^{w}(t_{j'-1})\cdot E_j\left\{\frac{d_{j'k}^{rw}}{Y_{j'k}}\right\}\right]\\
&= \sum_{j'=1}^{j}\left[S_k^{w}(t_{j'-1})\cdot\frac{\lambda_k^{rw}(t_{j'})\cdot Y_{j'k}}{Y_{j'k}}\right]\\
&= \sum_{j'=1}^{j}\lambda_k^{rw}(t_{j'})\cdot S_k^{w}(t_{j'-1})\\
&= I_k^{rw}(t_j).
\end{aligned}
$$

A.2 VARIANCE AND COVARIANCE OF THE ACIF

$$\begin{aligned} Var\left\{\hat{\lambda}_i^{rw}\hat{S}^w(t_{i-1})\right\} &= Var\left[E_{i-1}\left\{\hat{\lambda}_i^{rw}\hat{S}^w(t_{i-1})|\text{ information up to time } t_{i-1}\right\}\right] \\ &+ E\left[Var_{i-1}\left\{\hat{\lambda}_i^{rw}\hat{S}^w(t_{i-1})|\text{ information up to time } t_{i-1}\right\}\right] \\ &= Var\left[\hat{\lambda}_i^{rw}E_{i-1}\left\{\hat{S}^w(t_{i-1})\right\}\right] + E\left[(\hat{\lambda}_i^{rw})^2 Var_{i-1}\left\{\hat{S}^w(t_{i-1})\right\}\right] \\ &= \{S^w(t_{i-1})\}^2 \cdot Var(\hat{\lambda}_i^{rw}) + Var_{i-1}\left\{\hat{S}^w(t_{i-1})\right\} \cdot E(\hat{\lambda}_i^{rw})^2. \end{aligned}$$

If the probabilities of treatment, p_i, are known, then from Xie and Liu (2005) we can prove that

$$Var(\hat{\lambda}_i^{rw}) = \frac{\lambda_i^{rw}(1-\lambda_i^{rw})}{M_i}.$$

Moreover, if $\max_{i:T_i\geq t_j}(1/p_i)/\sum_{i:T_i\geq t_j} 1/p_i \to 0$, then the variance of the adjusted Kaplan-Meier function can be estimated and therefore

$$Var_{i-1}\left\{\hat{S}^w(t_{i-1})\right\} = \{S^w(t_{i-1})\}^2 \sum_{l=1}^{i-1} \frac{1-s^w(t_l)}{M_l s^w(t_l)}$$

and

$$E(\hat{\lambda}_i^{rw})^2 = Var(\hat{\lambda}_i^{rw}) + \left\{E(\hat{\lambda}_i^{rw})\right\}^2 = \frac{\lambda_i^{rw}(1-\lambda_i^{rw})}{M_i} + (\lambda_i^{rw})^2,$$

where $M_i = (\sum_{j:T_j\geq t_i} 1/p_j)^2/\sum_{j:T_j\geq t_i}(1/p_j)^2$ and $s^w(t_l) = S^w(t_l)/S^w(t_{l-1})$. With these estimates, the ACIF variance estimator has the form

$$\widehat{Var}\left\{\hat{\lambda}_i^{rw}\hat{S}^w(t_{i-1})\right\} = \left\{\hat{S}^w(t_{i-1})\right\}^2 (\hat{\lambda}_i^{rw})^2 \left[\frac{1-\hat{\lambda}_i^{rw}}{M_i\hat{\lambda}_i^{rw}} + \left\{\sum_{l=1}^{i-1}\frac{1-\hat{s}^w(t_l)}{M_l\hat{s}^w(t_l)}\right\}\left(\frac{1-\hat{\lambda}_i^{rw}}{M_i\hat{\lambda}_i^{rw}}+1\right)\right].$$

If we let $i < i'$, then the ACIF covariance estimator is

$$\begin{aligned} & Cov\left\{\hat{\lambda}_i^{rw}\hat{S}^w(t_{i-1}),\ \hat{\lambda}_{i'}^{rw}\hat{S}^w(t_{i'-1})\right\} \\ = \ & E\left\{\hat{\lambda}_i^{rw}\hat{S}^w(t_{i-1})\cdot\hat{\lambda}_{i'}^{rw}\hat{S}^w(t_{i'-1})\right\} - E\left\{\hat{\lambda}_i^{rw}\hat{S}^w(t_{i-1})\right\}E\left\{\hat{\lambda}_{i'}^{rw}\hat{S}^w(t_{i'-1})\right\} \\ = \ & \underbrace{E\left[\hat{\lambda}_i^{rw}\hat{\lambda}_{i'}^{rw}\hat{s}^w(t_i)\cdots\hat{s}^w(t_{i'-1})\cdot\left\{\hat{S}^w(t_{i-1})\right\}^2\right]}_{(1)} - \lambda_i^{rw}S^w(t_{i-1})\cdot\lambda_{i'}^{rw}S^w(t_{i'-1}). \end{aligned}$$

Part (1) can be further simplified as

$$
\begin{aligned}
& E\left\{E_{i-1}\left[\hat{\lambda}_i^{rw}\hat{\lambda}_{i'}^{rw}\hat{s}^w(t_i)\cdots\hat{s}^w(t_{i'-1})\left\{S^w(t_{i-1})\right\}^2 \mid \text{information up to time } t_{i-1}\right]\right\} \\
= \;& E\left\{\hat{\lambda}_i^{rw}\hat{\lambda}_{i'}^{rw}\hat{s}^w(t_i)\cdots\hat{s}^w(t_{i'-1})\cdot E_{i-1}\left[\left\{S^w(t_{i-1})\right\}^2\right]\right\} \\
= \;& E\left\{\hat{\lambda}_i^{rw}\hat{\lambda}_{i'}^{rw}\hat{s}^w(t_i)\cdots\hat{s}^w(t_{i'-1})\cdot\left[Var\left\{\hat{S}^w(t_{i-1})\right\}+E_{i-1}^2\left\{S^w(t_{i-1})\right\}\right]\right\} \\
= \;& E\left[\hat{\lambda}_i^{rw}\hat{\lambda}_{i'}^{rw}\hat{s}^w(t_i)\cdots\hat{s}^w(t_{i'-1})\cdot\left\{S_w(t_{i-1})\right\}^2\left\{\sum_{l=1}^{i-1}\frac{1-s^w(t_l)}{M_l s^w(t_l)}+1\right\}\right] \\
= \;& \lambda_{i'}^{rw}\{s^w(t_{i+1})\cdots s^w(t_{i'-1})\}S_w(t_{i-1})^2\{\sum_{l=1}^{i-1}\frac{1-s^w(t_l)}{M_l s^w(t_l)}+1\}\cdot\underbrace{E\{\hat{\lambda}_i^{rw}\cdot\hat{s}^w(t_i)\}}_{(2)}.
\end{aligned}
$$

If we split the total number of deaths at time t into death due to event r and death due to the non-events $\bar{r}$, $d_i^w = d_i^{rw} + d_i^{\bar{r}w}$, then part (2) becomes

$$
\begin{aligned}
& E\{\hat{\lambda}_i^{rw}\cdot\hat{s}^w(t_i)\} \\
= \;& E\{\frac{d_i^{rw}}{Y_i^w}(1-\frac{d_i^{rw}+d_i^{\bar{r}w}}{Y_i^w})\} \\
= \;& E(\hat{\lambda}_i^{rw}) - E\frac{(d_i^{rw})^2+d_i^{rw}\cdot d_i^{\bar{r}w}}{(Y_i^w)^2} \\
= \;& \lambda_i^{rw} - \frac{V(d_i^{rw})+E^2(d_i^{rw})+E(d_i^{rw})\cdot E(d_i^{\bar{r}w})}{(Y_i^w)^2} \\
= \;& \lambda_i^{rw} - \lambda_i^{rw}(\frac{1-\lambda_i^{rw}}{M_i}+\lambda_i^{rw}+\lambda_i^{\bar{r}w}) \\
= \;& \lambda_i^{rw}(1-\lambda_i^{w}-\frac{1-\lambda_i^{rw}}{M_i}) \\
= \;& \lambda_i^{rw}\{s^w(t_i)-\frac{1-\lambda_i^{rw}}{M_i}\} \\
= \;& \lambda_i^{rw}s^w(t_i)\{1-\frac{1-\lambda_i^{rw}}{M_i s^w(t_i)}\}.
\end{aligned}
$$

Thus, the covariance of the ACIF and its estimate have the form

$$
\begin{aligned}
& Cov\left\{\hat{\lambda}_i^{rw}\hat{S}^w(t_{i-1}),\ \hat{\lambda}_{i'}^{rw}\hat{S}^w(t_{i'-1})\right\} \\
= \;& \lambda_i^{rw}\lambda_{i'}^{rw}S^w(t_{i-1})S^w(t_{i'-1})[\{\sum_{l=1}^{i-1}\frac{1-s^w(t_l)}{M_l s^w(t_l)}+1\}\{1-\frac{1-\lambda_i^{rw}}{M_i s^w(t_i)}\}-1]
\end{aligned}
$$

and

$$\begin{aligned}
&\widehat{Cov}\left\{\hat{\lambda}_i^{rw}\hat{S}^w(t_{i-1}),\ \hat{\lambda}_{i'}^{rw}\hat{S}^w(t_{i'-1})\right\} \\
= \ &\hat{\lambda}_i^{rw}\hat{\lambda}_{i'}^{rw}\hat{S}^w(t_{i-1})\hat{S}^w(t_{i'-1})[\{\sum_{l=1}^{i-1}\frac{1-\hat{s}^w(t_l)}{M_l\hat{s}^w(t_l)}+1\}\{1-\frac{1-\hat{\lambda}_i^{rw}}{M_i\hat{s}^w(t_i)}\}-1],
\end{aligned}$$

respectively.

A.3 DISTRIBUTION OF THE TEST STATISTIC

Z_0 To show that the test statistic $Z_0 = WI/\sqrt{Var(WI)}$ has a standard normal distribution for large samples under the null hypothesis, we begin by showing that $\sqrt{n}\{\hat{I}^{rw}(t)-I^{rw}(t)\}(t \leq \tau)$ is asymptotically equivalent to a sum of i.i.d. mean zero random variables. We then utilize Theorem 4.1 of Pepe (1991) to construct the formula of the consistent variance estimator and the asymptotically normal test.

Here we try to extend the usual counting process technique to the setting of the weighted survival data. We define $N_{ir}^w(t) = w_i I[T_i \leq t, \delta_i = 1, \epsilon_i = r]$ and $N_r^w(t) = \sum_{i=1}^n N_i(t) = \sum_{t_i \leq t}^n \delta_i \cdot w_i \cdot I[\epsilon_i = r]$, where r denotes the cause type, i denotes as the ith individual. Let

$$M_r^{i\prime}(t) = N_r^{iw}(t) - \int_0^t w_i\lambda_r^w(s)ds$$

and we can show that $M_r^{i\prime}(t)$ is a martingale, that is,

$$\begin{aligned}
E\{dM_r^{i\prime}(t)|F_{t-}\} &= E\{[dN_r^{iw}(t) - d\{\int_0^t w_i\lambda_r^w(s)ds\}]|F_{t-}\} \\
&= E\{dN_r^{iw}(t)|F_{t-}\} - E\{w_i\lambda_r^w(t)dt|F_{t-}\} \\
&= 0,
\end{aligned}$$

where F_{t-} denotes the history, or filtration, at an instant just prior to t. The above expectation is equal to 0since $\lambda_r^w(t)$ has a fixed value given F_{t-}. Thus, $M_r'(t) = \sum_1^n M_r^{i\prime}$ is also a martingale, and

$$
\begin{aligned}
& \sqrt{n}\left\{\hat{I}^{rw}(t) - I^{rw}(t)\right\} \\
= \; & \sqrt{n}[\int_0^t \frac{\hat{S}^{w-}(s)}{Y_r^w(s)} I\{Y_r^w(s) > 0\} dN_r^w(s) - \int_0^t S^w(s)\lambda_r^w(s)ds] \\
= \; & \sqrt{n}\int_0^t \frac{\hat{S}^{w-}(s)}{Y_r^w(s)} I\{Y_r^w(s) > 0\} dM_r'(t) + \sqrt{n}\int_0^t \frac{\hat{S}^{w-}(s)}{Y_r^w(s)} I\{Y_r^w(s) > 0\} d\{\int_0^t Y_r^w(s)\lambda_r^w(s)ds\} \\
- \; & \sqrt{n}\int_0^t S^w(s)\lambda_r^w(s)ds \\
= \; & \sqrt{n}\int_0^t \frac{\hat{S}^{w-}(s)}{Y_r^w(s)} I\{Y_r^w(s) > 0\} dM_r'(t) + \sqrt{n}\int_0^t \hat{S}^w(s) I\{Y_r^w(s) > 0\}\lambda_r^w(s)ds - \sqrt{n}\int_0^t S^{w-}(s)\lambda_r^w(s)ds \\
= \; & \sqrt{n}\int_0^t \frac{\hat{S}^{w-}(s)}{Y_r^w(s)} I\{Y_r^w(s) > 0\} dM_r'(t) + \sqrt{n}\int_o^t \{\hat{S}^w(s) - S^w(s)\}\lambda_r^w(s)ds \\
- \; & \sqrt{n}\int_0^t \hat{S}^{w-}(s) I\{Y_r^w(s) = 0\}\lambda_r^w(s)ds.
\end{aligned}
$$

The third term converges to 0 in probability since $P\{Y_r^w(t) = 0\} \to 0$; the first term is asymptotically equivalent to $(\sqrt{n})^{-1}\sum_{i=1}^n \int_0^t \{\frac{S^w(s)}{Y_r^w(s)/n}\} dM_r^{i\prime}(s)$ and the second term is asymptotically equivalent to $(\sqrt{n})^{-1}\sum_{i=1}^n X_s^i(s)\lambda_r^w(s)ds$. Thus, we can show that $\sqrt{n}\{\hat{I}^{rw}(t) - I^{rw}(t)\}$ is asymptotically equivalent to $(\sqrt{n})^{-1}\sum_{i=1}^n X_{I_r^w}^i(t)$, where

$$
X_{I_r^w}^i(t) = \int_0^t \{\frac{S^w(s)}{Y_r^w(s)/n}\} dN_r^{iw}(s) - \int_0^t nS^w(s)\lambda_r^w(s)ds + \int_0^t X_S^i(s)\lambda_r^w(s)ds.
$$

Based on a Taylor series expansion, Corollary 2.1 of Pepe (1991) also holds for our ACIF. Therefore according to Theorem 4.1 of Pepe, we have:

$$
WI = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \cdot \int_0^\tau K(t)\left\{\hat{I}_1^{rw}(t) - \hat{I}_2^{rw}(t)\right\} dt \to^d N(0, \sigma_{WI}^2)
$$

and

$$
\hat{\sigma}_{WI}^2 = \sum_{j=1}^2 \frac{n^{3-j}}{(n^1 + n^2)n^j} \sum_{i=1}^{n^l} \{\int_0^\tau \hat{K}(u)\hat{X}_{I_r^w}^i(u)du\}^2 \to^P \sigma_{WI}^2.
$$

If $\int_0^\tau K(u)\{\hat{I}_1^{rw}(u) - \hat{I}_2^{rw}(u)\}du \neq 0$, then a test based on $WI/\sqrt{\hat{\sigma}_{WI}^2}$ and its asymptotic null distribution is consistent.

BIBLIOGRAPHY

[1] Aalen O. Nonparametric estimation of partial transition probabilities in multiple decrement models. *Annals of Statistics* 1978; **6**:534–545.

[2] Amato DA. A generalized Kaplan-Meier estimator for heterogeneous populations. *Communication in Statistics-Theory and Methods* 1988; **17**:263–286.

[3] Braun TM, Yuan Z. Comparing the small sample performance of several variance estimators under competing risks. *Statistics in Medicine* 2007; **26**:1170–1180.

[4] Cheng SC, Fine JP, Wei LJ. Prediction of cumulative incidence function under the proportional hazards model. *Biometrics* 1998; **54**:219–228.

[5] Cupples LA, Gagnon DR, Ramaswamy R, D'Agostino RB. Age-adjusted survival curves with application in the Framingham study. *Statistics in Medicine* 1995; **14**:1731–1744.

[6] Dawson R, Lavori RW. Using inverse weighting and predictive inference to estimate the effects of time-varying treatments on the discrete-time hazard. *Statistics in Medicine* 2002; **21**:1641–1661.

[7] Dinse GE, Larson MG. A note on semi-Markov models for partially censored data. *Biometrika* 1986; **73**:379–386.

[8] Faucett CL, Schenker N, Taylor JMG. Survival analysis using auxiliary variables via multiple imputation, with application to AIDS clinical trial data. *Biometrics* 2002,**58**:37–47.

[9] Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association* 1999; **94**:496–509.

[10] Finkelstein DM, Schoenfeld DA. Analysing survival in the presence of an auxiliary variable. *Statistics in Medicine* 1994; **13**:1747–1754.

[11] Fleming TR, Prentice RL, Pepe MS, Glidden D. Surrogate and auxiliary endpoints in clinical trials, with potential applications in cancer and AIDS research. *Statistics in Medicine* 1994; **13**:955–968.

[12] Gaynor JJ, Feuer EJ, Tan CC, Wu DH, Little CR, Straus DJ, Clarkson BD, Brennan MF. On the use of cause-specific failure and conditional failure probabilities: examples from clinical oncology data. *Journal of the American Statistical Association* 1993; **88**:400–409.

[13] Gooley TA, Leisenring W, Crowley J, Storer BE. Estimation of failure probabilities in the presence of competing risks: new presentations of old estimators. *Statistics in Medicine* 1999; **18**:695–706.

[14] Gray R. A kernel method for incorporating information on disease progression in the analysis of survival. *Biometrika* 1992; **81**:527–539.

[15] Gray R. A class of k-sample tests for comparing the cumulative incidence of a competing risk. *Annals of Statistics* 1988; **16**:1141–1154.

[16] Grunkemeier GL, Jin R, Eijkemans MJC, Takkenberg JJM. Actual and actuarial probabilities of competing risks: apples and lemons. *The Annals of Thoracic Surgery* 2007; **83**:1586–1592.

[17] Hankey BF, Myers MH. Evaluating differences in survival between two groups of patients. *Journal of the Chronic Disease* 1971; **24**:523–531.

[18] Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* 1958; **53**:457–481.

[19] Klein JP, Moeschberger ML. Bounds on net probabilities for dependent competing risks. *Biometrics* 1988; **44**:529–538.

[20] Klein JP, Moeschberger ML. *Survival Analysis: Techniques for Censored and Truncated Data*, 2nd edition. Springer: New York, 2003.

[21] Lagakos SW, Williams JS. Models for censored survival analysis: a cone class of variable-sum models. *Biometrika* 1978; **65**:181-189.

[22] Lin DY. Non-parametric inference for cumulative incidence functions in competing risks studies. *Statistics in Medicine* 1997; **16**:901–910.

[23] Link WA, A model for informative censoring. *Journal of the American Statistical Association* 1989; **84**:749–752.

[24] Little RJA. Survey nonresponse adjustments. *International Statistical Review* 1986; **54**:139–157.

[25] Lunceford JK, Davidian M. Stratification and weighting via the propensity score in estimation of causal treatment effects: A comparative study. *Statistics in Medicine* 2004; **23**:2937–2960.

[26] Lunn M, McNeil D. Applying Cox regression to competing risks. *Biometrics* 1995; **51**:524–532.

[27] Malani HM. Modification of the re-distribution to the right algorithm using disease markers. *Biometrika* 1995; **82**:515–526.

[28] Moeschberger ML, Klein JP. Statistical methods for dependent competing risks. *Lifetime data analysis* 1995; **1**:195–204.

[29] Murray S, Tsiatis AA. Non-parametric survival estimation using prognostic longitudinal covariates. *Biometrics* 1996; **52**:137–151.

[30] Nieto FJ, Coresh J. Adjusting survival curves for confounders: a review and a new method. *American Journal of Epidemiology* 1996; **143**:1059–1068.

[31] Organ Procurement and Transplantation Network. Available at www.optn.org. Accessed November, 2004.

[32] Pepe MS. Inference for events with dependent risks in multiple endpoint studies. *Journal of the American Statistical Association* 1991; **86**:770–778.

[33] Pepe MS, Fleming TR. Weighted Kaplan-Meier statistics - a class of distance tests for censored survival data. *Biometrics* 1989; **45**:497–507.

[34] Pepe MS, Fleming TR. Weighted Kaplan-Meier statistics: Large sample and optimality considerations. *Journal of the Royal Statistical Society, Series B* 1991; **53**:341–352.

[35] Pepe MS, Mori M. Kaplan-Meier, marginal or conditional probability curves in summarizing competing risks failure time data? *Statistics in Medicine* 1993; **12**:737–751.

[36] Peterson AV, Bounds for a joint distribution function with fixed subdistribution functions: Application to competing risks. *Proceedings of the National Academy of Science* 1976; **73**:11–13.

[37] Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weight (IPCW) log-rank tests. *Biometrics* 2000; **56**:779–788.

[38] Robins JM. Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers. *In proceedings of the Biopharmaceutical section, American Statistical Association*, 24-33. Alexandra, Virginia: American Statistical Association, 1993.

[39] Robins JM, Rotnitzky A. Recovery of Information and Adjustment for Dependent Censoring Using Surrogate Markers. *Aids and Epidemiology-Methodological Issues* 1992; **Boston:Birkhauser**:24-33.

[40] Rosenbaum PR, Rubin DB. Constructing a control group using multivariate matched sampling incorporating the propensity score. *The American Statistician* 1985; **39**:33–38.

[41] Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983; **70**:41–55.

[42] Rotnitzky A, Farall A, Bergesio A, Scharfstein D. Analysis of failure time data under competing censoring mechanisms. *Journal of the Royal Statistical Society, Series B* 2007; **69**:307–327.

[43] Satten GA, Datta S. The Kaplan-Meier Estimator as an Inverse-Probability-of-Censoring Weighted Average. *The American Statistician* 2001; **55**:207–210.

[44] Satten GA, Datta S, Robins JM. Estimating the marginal survival function in the presence of time dependent covariates. *Statistics & Probability Letters* 2001; **54**:397–403.

[45] Satagopan JM, Ben-Porat L, Berwick M, Robson M, Kutler D, Auerbach AD. A note on competing risks in survival data analysis. *British Journal of Cancer* 2004; **91**:1229–1235.

[46] Slud EV, Rubinstein LV. Dependent competing risks and summary survival curves. *Biometrika* 1983; **70**:643–649.

[47] Scharfstein DO, Robins JM. Estimation of the failure time distribution in the presence of informative censoring. *Biometrika* 2002; **89**:617–634.

[48] Wang CY, Wang S, Zhao LP, Ou ST. Weighted semiparametric estimation in regression analysis with missing covariate data. *Journal of the American Statistical Association* 1997; **92**:512–525.

[49] Wei LJ, Lin DY, Weissfeld LA. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association* 1989; **84**:1065–1073.

[50] Williams JS, Lagakos SW. Models for censored survival analysis: constant-sum and variable-sum models. *Biometrika* 1977; **64**:215–224.

[51] Winnett A, Sasieni P. Adjusted Nelson-Aalen estimates with retrospective matching. *Journal of the American Statistical Association* 2002; **97**:245–256.

[52] Xie J, Liu C. Adjusted Kaplan-Meier estimator and log-rank test with inverse probability of treatment weighting for survival data. *Statistics in Medicine* 2005; **24**:3089–3110.

[53] Zheng M, Klein JP. A self-consistent estimator of marginal survival functions based on dependent competing risk data and an assumed copula. *Communications in Statistics A* 1994; **23**:2299–2311.

[54] Zheng M, Klein JP. Estimates of marginal survival for dependent competing risks based on an assumed copula. *Biometrika* 1995; **82**:127–138.